University of Alberta

Transmission of H.263 Video Over ATM Networks

by

Robert Shaffer

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of **Master of Science**.

Department of Computing Science

Edmonton, Alberta
Spring 1999

National Library of Canada
Acquisitions and Bibliographic Services
395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale du Canada
Acquisitions et services bibliographiques
395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-40108-1

Canada

University of Alberta

Library Release Form

Name of Author: Robert Shaffer

Title of Thesis: Transmission of H.263 Video Over ATM Networks

Degree: Master of Science

Year this Degree Granted: 1999

Permission is hereby granted to the University of Alberta Library to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves all other publication and other rights in association with the copyright in the thesis, and except as hereinbefore provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatever without the author's prior written permission.

Robert Shaffer

Robert Shaffer
P.O. Box 5462
Westlock, Alberta
Canada, T7P 2P5

Date: Nov. 02, 1998

University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled **Transmission of H.263 Video Over ATM Networks** submitted by Robert Shaffer in partial fulfillment of the requirements for the degree of **Master of Science.**

. . . . . J. Harms . . . . .

Janelle Harms

. . . . . .

Anup Basu

. . . . . .

Curtis Hrischuk

. . . . . .

Ursula Maydell

Date: Nov. 02, 1998

Abstract

In this thesis we address issues relating to the transmission of H.263 coded video over Asynchronous Transfer Mode (ATM). We simulate the transmission of Variable Bit Rate (VBR) video over a Constant Bit Rate (CBR) ATM connection, avoiding the loss of data associated with VBR connections when congestion occurs. First, we show that transmitting VBR coded H.263 video over a CBR connection without taking buffer overflow into account can cause significant degradation of the video quality. We then discuss the performance of the two video encoders that we introduce, they are: *black box* and *dynamic*. Using the dynamic video encoder we analyze the performance of two video coding strategies; the first strategy uses I-frames dispersed among P-frames, while the second strategy, partial encoding, also codes part of the (fixed) I-frames along with the (predictive) P-frames. It is demonstrated that partial encoding of fixed frames gives a better perceptual image quality and network performance of the two proposed methods.

Acknowledgements

I would like to thank my supervisors Janelle Harms and Anup Basu for their encouragement and insight into the problems encountered. The suggestions Janelle provided related to communication networks and feedback pertaining to the thesis document are greatly appreciated. Similarly, the discussions with Anup regarding the challenges pertaining to image compression were important in formulating the problem addressed in this thesis. In addition, the support and direction he provided are truly appreciated.

The H.263 source code developed by Telenor Research and Development of Norway, is a very important part of this thesis. The preliminary work done by Candy Pang as part of her course project, included documenting and evaluating the code which was helpful as the documentation provided in the source code was very scant. Lijun Yin provided some very important input during development of the thesis related to work he had done with the H.261 video compression standard, a predecessor to H.263.

The input provided by members of our Networks Research group at weekly meetings was also important to this study. In particular, the assistance provided by Dr. Ioanis Nikolaidis, and Wladyslaw Olesinski was very helpful in formulating a part of the performance assessment criteria used in this work.

Finally, I would especially like to thank my family for the encouragement and suggestions they provided during the course of my studies. Their support contributed immensely to the gratification I received while a graduate student at the University of Alberta.

Robert Shaffer

November 02, 1998

Contents

1 Introduction 1
- 1.1 Problem Description . . . 3
- 1.2 Overview of Our System and the Assumptions Pertaining to This Work . . 4
- 1.3 Thesis Overview . . . 6

2 H.263 Video Compression 8
- 2.1 Overview of the H.263 Video Compression Standard . . . 8
- 2.2 CBR and VBR Video Encoding . . . 12
  - 2.2.1 A Comparison of CBR and VBR Encoding . . . 12
  - 2.2.2 Evaluation of the H.263 Constant Bit Rate Option . . . 13

3 Background Networking Information 25
- 3.1 A Networking Overview of CBR and VBR Traffic . . . 25
- 3.2 Previous Work . . . 27
- 3.3 Asynchronous Transfer Mode (ATM) . . . 29

4 Design of our Video Encoder Algorithm 34
- 4.1 A Black Box Video Encoder . . . 34
  - 4.1.1 An Overview of the Black Box Encoding Algorithm . . . 34
  - 4.1.2 Implementation of the Black Box Video Encoder . . . 36
- 4.2 A Dynamic Video Encoder . . . 37
  - 4.2.1 An Overview of the Dynamic Encoding Algorithm . . . 37
  - 4.2.2 Implementation of the Dynamic Video Encoder . . . 38

5 Overview of Our Simulation Model and Assessment Criteria 40
- 5.1 Simulation Design . . . 40
  - 5.1.1 Simulation Model . . . 40
  - 5.1.2 Simulation Events . . . 41
  - 5.1.3 Implementation . . . 42
  - 5.1.4 Simplifications . . . 43
  - 5.1.5 Validation . . . 44
- 5.2 Assessment Criteria . . . 44
  - 5.2.1 Video Quality Assessment Criteria . . . 44
  - 5.2.2 Network Assessment Criteria . . . 49
- 5.3 Selection of Test Video Sequences . . . 50

6 Experiments Using Intra-encoding at the Frame Level 55
6.1 Performance When No Frames are Discarded 55
6.1.1 Limited CBR Connection Capacity, and Unlimited Buffering (Test 1) 56
6.1.2 No Buffering and Increased Capacity of the CBR Connection (Test 2) 57
6.2 Performance When Frames are Discarded Due to Buffer Overflow 58
6.2.1 Frame Discard Algorithms 58
6.2.2 Computing the CBR Connection Capacities and the Buffer Sizes . . 59
6.2.3 Results When a Black Box Video Encoder is Used and Buffer Overflow is Experienced (Test 3) 61
6.2.4 Results When a Dynamic Video Encoder is Used and Buffer Overflow is Experienced (Test 4) 64
6.2.5 Performance Comparison of Black Box and Dynamic Video Encoding 66
6.2.6 Discarding the Smallest and Largest Inter-encoded Frames (Test 5) . 69

7 Experiments Using Intra-encoding at the Macroblock Level 73
7.1 Macroblock Intra-encoding Algorithm 74
7.2 Performance When the Frequency of Macroblocks Intra-encoded is Varied (Test 6) 75
7.3 Various CBR Connection Capacities with Unlimited Buffering (Test 7) . . . 77
7.3.1 Computing the CBR Connection Capacities 78
7.3.2 Foreman and Miss America Results When Intra-encoded at the Macroblock Level 78
7.4 Performance Comparison of Intra-encoding at the Frame and Macroblock Levels 82
7.5 Results When the Carphone, Claire, and Salesman Video Sequences are Intra-encoded at the Macroblock Level 84
7.6 Summary of Results (Intra-encoding at the Macroblock Level) 86

8 Conclusion and Future Work 88
8.1 Conclusion 88
8.2 Future Work 90

Bibliography 92

A Glossary of Terms 95

B Glossary of Acronyms 97

List of Figures

1.1 Overview of the system components used in our experiments. 5

2.1 The H.263 hierarchical structure [35]. 11
2.2 The numbering sequence of the 99 macroblocks found in an H.263 video frame when QCIF is used. 11
2.3 A flowchart illustrating the H.263 CBR encoding algorithm. 14
2.4 Miss America frame sizes using VBR encoding. 16
2.5 Miss America frame sizes when CBR encoding is used, (encoding bit rate = mean VBR encoding bit rate). 17
2.6 Miss America quantization parameters when CBR encoding is used, (encoding bit rate = mean VBR encoding bit rate) 18
2.7 Miss America frame sizes when CBR encoding is used, (encoding bit rate = mean VBR encoding bit rate plus 10 percent) 18
2.8 Miss America quantization parameters when CBR encoding is used, (encoding bit rate = mean VBR encoding bit rate plus 10 percent) 19
2.9 Miss America frame sizes when CBR encoding is used, (encoding bit rate = mean VBR encoding bit rate plus 15 percent) 20
2.10 Miss America quantization parameters when CBR encoding is used, (encoding bit rate = mean VBR encoding bit rate plus 15 percent) 20
2.11 Miss America frame sizes when CBR encoding is used, (encoding bit rate = mean VBR encoding bit rate plus 80 percent) 21
2.12 Miss America quantization parameters when CBR encoding is used, (encoding bit rate = mean VBR encoding bit rate plus 80 percent) 21
2.13 Miss America frame sizes when CBR encoding is used, (encoding bit rate = mean VBR encoding bit rate plus 1000 percent) 22
2.14 Miss America quantization parameters when CBR encoding is used, (encoding bit rate = mean VBR encoding bit rate plus 1000 percent) 22
2.15 Miss America frame sizes when CBR encoding is used, (encoding bit rate = mean VBR encoding bit rate plus 3000 percent) 23
2.16 Miss America quantization parameters when CBR encoding is used, (encoding bit rate = mean VBR encoding bit rate plus 3000 percent) 24

3.1 The leaky bucket algorithm [33]. 27
3.2 The ATM cell header formats [33]. 30
3.3 The ATM reference model [33]. 31

4.1 The H.263 video encoder as a black box. 35
4.2 Encoding scheme when using a black box video encoder. 36

4.3 A dynamic H.263 video encoder. 37
4.4 Encoding scheme when using a dynamic video encoder. 38

5.1 Overview of the system components used in our experiments. 41
5.2 Development process for the video quality assessment algorithm [37]. 46
5.3 A frame in the Carphone video sequence. 51
5.4 A frame in the Claire video sequence. 52
5.5 A frame in the Foreman video sequence. 52
5.6 A frame in the Miss America video sequence. 52
5.7 A frame in the Salesman video sequence. 53

6.1 Video quality of the Foreman video sequence when a black box encoder is used and the small- and large-sized frames are discarded. 63
6.2 Video quality of the Miss America video sequence when a black box encoder is used and the small- and large-sized frames are discarded. 65
6.3 Video quality of the Foreman video sequence using both a black box and dynamic video encoder when the small- or large-sized frames are discarded. 68
6.4 Video quality of the Miss America video sequence using both a black box and dynamic video encoder when the small- or large-sized frames are discarded. 70

7.1 Resulting jitter values for the Miss America and Foreman video sequences using intra-encoding at the macroblock level. 81
7.2 Maximum buffer requirements for the Miss America and Foreman video sequences. 82

List of Tables

2.1 A comparison of the frames sizes when CBR encoding is used to compress the Miss America video sequence. 15

5.1 Spatial and temporal information present in the video sequences. 51
5.2 Frame sizes in the Foreman and Miss America video sequences. 53

6.1 Test 1 results, limited CBR connection capacity and unlimited buffering. . . 56
6.2 Test 2 results, no buffering and the CBR connection capacity is increased to the point where each frame is transmitted in a single frame period. 57
6.3 Test 3 results, Foreman video sequence when various CBR connection capacities are used with a black box encoder. 62
6.4 Test 3 results, Miss America video sequence when various CBR connection capacities are used with a black box video encoder. 64
6.5 Test 4 results, Foreman video sequence when various CBR connection capacities are used with a dynamic video encoder and small-sized frames are discarded. 66
6.6 Test 4 results, Miss America video sequence when various CBR connection capacities are used with a dynamic video encoder and small-sized frames are discarded. 66
6.7 Performance comparison of the black box and dynamic video encoders using the Foreman video sequence. 67
6.8 Performance comparison of the black box and dynamic video encoders using the Miss America video sequence. 69
6.9 Test 5 results, when only the smallest and largest P-frames are discarded. . 70
6.10 Video quality analysis when the smallest and largest P-frames are discarded. 71

7.1 Intra-encoding the Foreman video sequence at the macroblock level. 76
7.2 Intra-encoding the Miss America video sequence at the macroblock level. . . 77
7.3 Test 7 results, intra-encoding at the macroblock level using various link capacities and unlimited buffering for the Foreman video sequence. 79
7.4 Test 7 results, intra-encoding at the macroblock level using various link capacities and unlimited buffering for the Miss America video sequence. . . . 79
7.5 Frame sizes in the Foreman and Miss America video sequences when a single macroblock is intra-encoded in each P-frame. 81
7.6 Results obtained using the frame and macroblock level intra-encoding of the Foreman video sequence. 83
7.7 Results obtained using the frame and macroblock level intra-encoding of the Miss America video sequence. 84

7.8 A summary of the results obtained for the Carphone, Claire, and Salesman video sequences using various link capacities and unlimited buffering. 85
7.9 Summary of the frame sizes pertaining to the five sequences we tested when a single macroblock is intra-encoded in each P-frame. 86

Chapter 1

Introduction

A vast number of applications today require video transmission, and utilize larger amounts of bandwidths than in the past. Two main reasons for the increasing demand of video transmission include the abundance of personal computers and workstations and the increasing communication needs of the information society. Video is the displaying of frames (scenes) on a computer or television screen.

Applications motivating research in the area of low bit rate video encoding include: video games, surveillance, communication aids for deaf people and videophones [17]. Although bandwidth has recently become more plentiful, the low bit rate constraints imposed by the Public Switched Telephone Network (PSTN) connecting users to the high speed networks are still present. Therefore, there is a need to compress the large amount of visual information present in video sequences for transmission over a network mainly carrying voice traffic. Furthermore, broadband networks are expected to carry a wide variety of traffic types, *i.e.*, data, voice, or video, it is expected that video traffic due to its larger bandwidth requirements will dominate the usage of future communication networks [19]. Therefore, there is a requirement to reduce the number of bits produced by video sources. In order to reduce the number of bits generated by the video streams for transmission over the communication network, video compression is used.

Video compression is generally lossy which implies that the original video quality cannot be obtained after decompression, thus degradation of the video quality results. Lossy compression is attractive however, since it increases the degree of compression that can be achieved, decreasing the number of bits found in the compressed bit stream. However, the level of video quality required may be application dependent. For example, the quality of a video surveillance recording should allow law enforcement agencies to identify distinguishing features of individuals involved in criminal activity while applications such as a video

conference can tolerate greater losses and thus achieve greater compression. If information is lost and the resulting video quality deteriorates beyond what the user deems acceptable, the network resources and intended benefits of the video service are wasted.

There are a number of image compression standards in place which include Joint Photographic Experts Group (JPEG), Motion Picture Experts Group (MPEG) and those specified by the International Telecommunication Union, H.261 and H.263. JPEG compression is intended for still images, and research focused on improving this standard can be found in [38]. MPEG is comprised of a variety of standards (MPEG 1, 2, 4) intended for compressing video. MPEG differs from JPEG because MPEG also uses compression to exploit temporal redundancies between adjacent frames; adjacent frames with little or no motion have large amounts of temporal redundancies. Research discussing the transmission of MPEG over Asynchronous Transfer Mode (ATM) networks can be found in [12] and [10]. H.263's predecessor H.261 (discussed in [20]) is intended for bit rates in the range of 64 kbps to 2 Mbps, while H.263 (described in [11]) is designed for low bit rate video compression, rates less than 64 kbps. Research focused on improving the robustness of H.263 in lossy environments such as the Internet is discussed by Willebeek-LeMair and Shae in [39]. They use an algorithm (intended for non-real-time encoding) that selects the macroblocks for intra-encoding based on their impact on later frames. Image compression uses a hierarchy to achieve greater compression, macroblocks are a level in this hierarchy and represent areas of size 16x16 image pixels. Those macroblocks having the most impact are intra-encoded to reduce the error propagation.

The amount of spatial and temporal information present in video sequences results in a high variability among the compressed frame sizes. Transmission of these frames over a network results in Variable Bit Rate (VBR) traffic which may lead to low utilization of the communication network, and/or bit losses when the available network connection capacities are exceeded. Hence, the transmission of compressed video frames is another active area of research. A video encoder may modify encoding parameters to reduce the variability of the frame sizes, this is Constant Bit Rate (CBR) encoding. The research performed by Baldi and Ofek indicates that the delays associated with CBR encoding are unacceptable [2]. VBR image compression is preferred over CBR encoding because it provides a more uniform image quality since the amount of information in the video sequence changes with time.

A large amount of research is devoted to devising algorithms that further reduce the encoded bit rate and improve the resulting video quality. Methods of achieving this goal

include optimizing the segmentation and motion estimation [14], [36], [40], and optimizing the allocation of the bits used to encode the compressed video stream [5], [15], and [30] to further enhance the performance of existing standards.

In addition to CBR and VBR video encoding, there also exist CBR and VBR network traffic which have their own advantages and disadvantages. ATM networks provide both CBR and VBR Quality of Service (QoS). A CBR ATM QoS provides dedicated connections to the users and therefore does not experience congestion within the network. If the peak rate of the CBR connection is not exceeded, the possibility of the user's data being discarded due to congestion is eliminated. However, because it is a dedicated connection, the network provider may charge the user for the capacity reserved instead of the capacity used; therefore the user will prefer to obtain a high utilization of the connection. The utilization of a connection, is a percentage computed from the ratio of the user's transmitted data divided by the capacity available. For example, if a 50 kbps connection is available, and the user on average transmits 25 kbps, the utilization of the connection is 50%.

ATM also offers VBR service, the capacity of the connection may be specified by the user's average bit rate. Unlike CBR service, a VBR connection is not dedicated to a single user, which allows a greater number of users to use the network. However, if the aggregate capacity of the users' data exceeds the physical capacity of the network, congestion arises, and the data of one of more users is discarded.

1.1 Problem Description

In this thesis we investigate the transmission of H.263 compressed video streams over ATM networks; the application we have chosen is video conferencing. Video conferencing applications such as Internet chat and distance education have become abundant and require real-time video communication.

The H.263 video compression standard is intended to provide video telephony service using bit rates below 64 kbps [21]. Video compression is achieved through the exploitation of either spatial or temporal redundancies found in the video sequence. Spatial redundancies include those found in a single video frame, such as a frame where the background is a wall and all the pixels corresponding to the background have the same setting. Temporal redundancies correspond to the overlap of information between adjacent frames. A video sequence containing very little motion would achieve greater amounts of compression due to the temporal redundancies that exist since the adjacent frames would be very similar.

Since the amount of compression achieved is dependent upon the amount of spatial and

temporal information present in each frame, the resulting compressed bit stream may be highly variable (variable bit rate encoding). As stated before, VBR encoding is preferred in the field of image compression since the resulting scene quality remains constant. CBR video encoding requires modifying certain parameters which adjust the amount of detail encoded in an attempt to keep the resulting number of bits for each frame nearly constant. This results in a varying video quality which the user may find annoying or possibly unacceptable.

A CBR network connection is preferred for video conferencing and other real-time applications since there are no queuing delays within the network because the user has a dedicated network connection between the parties involved in the conference. This reduces the inter-frame display times known as jitter [2]. However, since CBR service provides a dedicated connection to the user, periods when the amount of compressed video frame data present is low, the connection is under utilized. Similarly, during periods of high motion for example that generate a large number of compressed bits, the CBR connection capacity may be exceeded causing information to be lost.

The transmission of a bursty video bit stream over a communication network adversely affects the following constraints: queuing delays, network utilization, jitter and the resulting video quality. In this thesis we investigate the ability to transmit a H.263 VBR encoded bit stream over a CBR network connection in order to achieve the advantages of VBR encoding and overcome the disadvantages pertaining to the use of a CBR network connection. The goal of our work is to obtain a high utilization of the CBR connection used to transmit the VBR encoded H.263 video while maintaining an acceptable level of jitter, less than or equal to 4 msec.

1.2 Overview of Our System and the Assumptions Pertaining to This Work

This section briefly outlines the design of our simulation and the assumptions pertaining to our work. Figure 1.1 illustrates the main components of the system used in our experiments. The input to the video encoder is a file in QCIF (Quarter Common Intermediate Format) containing the Y-component (luminance) followed by the U and V components (containing the color information) for each frame in the video sequence. QCIF specifies the resolution of the frames; the QCIF resolution is 144 lines by 176 pixels/line [11]. We have five standard video sequences stored in this format, each consists of 150 frames that are used in our experiments with a frame refresh rate of 30 frames/second. The H.263 video encoder and decoder source code that we use is version 2.0 developed by Telenor R&D, Norway;

the source code was obtained off the world wide web [34]. Following the H.263 compression standard, the video encoder compresses each frame of the video sequence which is then added to the buffer used to smooth the VBR video data stream. Data is then removed from the buffer at a constant rate determined by the capacity of the simulated CBR ATM connection and then transmitted to the decoder. Because the video encoder does not run in real-time, the output from the decoder is written to a file that the video decoder will uncompresses to playback the video sequence on the computer monitor.

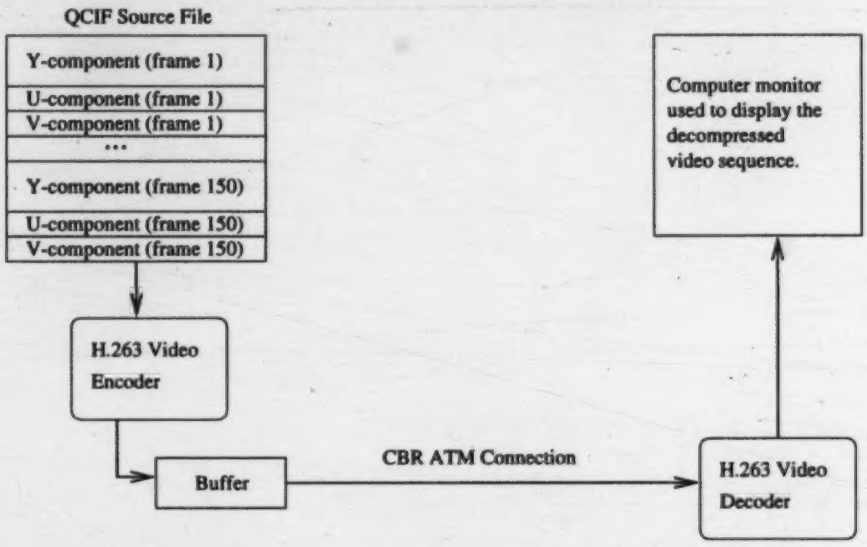

Figure 1.1: Overview of the system components used in our experiments.

The documentation provided with the video encoder source code was very scant and mostly confined to the function headers. Therefore, in order to facilitate the modifications required, such as, adding the video quality assessment algorithm used in our experiments, considerable time was spent documenting the code. The video quality assessment algorithm designed by Webster *et al.* [37], is an objective evaluation of the decompressed video quality (discussed in Section 5.2.1). The familiarity gained when documenting the H.263 source code helped immensely with the implementation and debugging of our modifications.

There are two main assumptions made in this work: frames are added to the buffer as complete units and losses are not experienced in the simulated CBR ATM connection. The H.263 video encoding standard is intended for real-time applications. However, since the source code we have does not run in real-time, the processing time required to encode each frame in a real-time scenario is not known. Therefore, entire compressed H.263 frames are

added to the buffer every 1/30th of a second (in simulation time) instead of as a stream of data. We use a simulated CBR ATM connection to transmit the data between the encoder and decoder. Since a CBR connection provides the user with a dedicated connection, we do not model background traffic. In addition, because the probability of a bit error in a fiber optic network is about 10^{-9} [10], we do not model bit errors within the network.

1.3 Thesis Overview

Given the tradeoffs between CBR and VBR traffic from the video compression and communications network viewpoints, we propose a method which will facilitate the transmission of VBR video encoding over a CBR network connection. The H.263 standard recommends that macroblocks be intra-encoded at least once every 132 times to reduce the propagation of errors[11]. Tests 1-5 use intra-encoding at the frame level (discussed in Chapter 6) causing frame numbers 1 and 100 to be intra-encoded. In all of our experiments the first frame is intra-encoded. Tests 7 and 8 use intra-encoding at the macroblock level (described in Chapter 7) where a single macroblock is intra-encoded in each frame following the first frame. A macroblock corresponds to an area of size 16x16 pixels in a video frame. A description of the experiments we perform are described below.

- Limited CBR Connection Capacity and Unlimited Buffering (Test 1):

  The capacity of the CBR connection is set equal to the mean bit rate, the buffer requirements and jitter (standard deviation of the inter-frame display time) are monitored.

- No Buffering and Increased Capacity of the CBR Connection (Test 2):

  The capacity is increased to the point where congestion at the buffer is alleviated since each frame can be transmitted in 1/30th of a second; we measure how much the CBR connection capacity must be increased over the mean bit rate and the utilization of the CBR connection.

- A Black Box Video Encoder is Used and Buffer Overflow is Experienced (Test 3):

  We monitor the performance of a black box video encoder (described in Section 4.1) based on the video quality produced when buffer overflow is experienced; the CBR connection capacity is varied from the mean bit rate to the rate when the largest frame can be transmitted in 1/30th of a second.

- A Dynamic Video Encoder is Used and Buffer Overflow is Experienced (Test 4):

  Test 4 is similar to Test 3, except that we use a dynamic video encoder (described in Section 4.2).

- Discarding the Smallest and Largest Inter-Encoded Frames (Test 5):

  Test 5 discards the smallest, and then the largest inter-encoded frames in order to evaluate whether small- or large-sized frames contribute more to the video quality. A similar number of bits are transmitted when both the smallest and largest frame sizes are discarded to provide a fair evaluation.

- Performance Evaluation When the Frequency of Macroblocks Intra-encoded is Varied (Test 6):

  Test 6 is used to evaluate the performance when the number of macroblocks intra-encoded in each frame is varied based on the video quality produced and the number of bits generated; when the number of macroblocks intra-encoded is increased, the compressed frame size increases.

- Various CBR Connection Capacities with Unlimited Buffering (Test 7):

  Test 7 uses intra-encoding at the macroblock level and monitors the utilization of the CBR connection and the corresponding jitter when the CBR connection capacity is increased from the mean bit rate to a capacity where the jitter is reduced to an acceptable level.

The layout of this thesis is as follows; Chapter 2 includes an overview of the H.263 video compression standard, an introduction to CBR and VBR video encoding, and a performance evaluation of the H.263 video encoder's CBR encoding option. In Chapter 3, we describe background networking information which includes an overview of CBR and VBR traffic, followed by an overview of ATM networks. Chapter 4 provides a description of the black box and dynamic video encoder algorithms we developed, followed by a description of our simulation model and assessment criteria in Chapter 5. A description and analysis of five standard video sequences available for testing is also provided in Chapter 5. In Chapter 6, we discuss the results obtained when intra-encoding is performed at the frame level and evaluate the performance of the black box and dynamic video encoders. This is followed by a discussion of our results obtained when intra-encoding is performed at the macroblock level in Chapter 7. In Chapter 8, we discuss the conclusions drawn from our work and outline the direction of our future work.

Chapter 2

H.263 Video Compression

The following section provides an overview of the H.263 video compression standard and a description of the video applications that H.263 and other video encoding standards such as MPEG are intended. In Section 2.2 we summarize the tradeoffs between CBR and VBR video encoding, and conclude with an evaluation of the H.263 video encoder's CBR encoding option.

2.1 Overview of the H.263 Video Compression Standard

This section describes the H.263 video compression standard, and includes: a description of the various video compression frame types, a discussion of how video compression standards such as MPEG relate to H.263, and a description of the H.263 video encoding hierarchy. Video compression is achieved by exploiting the spatial and temporal redundancies present in a video sequence. The former corresponds to redundancies found in single frames, a frame containing very little detail has large spatial redundancy. Temporal redundancies pertain to redundancies between adjacent frames; large temporal redundancies are found in video frames containing little or no motion.

I- (Intra) frames, P- (Predictive) frames and B- (Bidirectional) frames are used in the Motion Picture Experts Group (MPEG) standard, and are also part of the H.261 and H.263 standards. I-frames exploit only spatial redundancies since they achieve video compression without reference to other frames. P-frames and B-frames exploit both the spatial and temporal redundancies present to achieve greater compression gains.

Although I-frames generally do not achieve the compression gains of either P-frames or B-frames they tend to be more robust in environments where the transmission medium is unreliable, as may be encountered on the Internet. P-frames, or predicted frames, reference the previously encoded frame in order to exploit temporal redundancies in addition to

the spatial redundancies exploited when intra-encoding is used. B-frames, or bidirectional frames are similar to P-frames, except that they reference both the previously decoded P-frame and the P-frame currently being decoded. Depending on the amount of motion present, the inter-encoded frames (P- and B-frames) may be much smaller (typically 2 to 4 times) than the intra-encoded frames [2]. For example if the scene changes are very small, the temporal redundancy will be very large, facilitating larger compression gains.

H.263 is a video compression standard defined by the Telecommunications standardization sector of the International Telecommunication Union (ITU-T). Its predecessor, H.261, was designed for the transmission of compressed video over the Public Switched Telephone Network (PSTN) using bit rates in the order of 64 kbps to 2 Mbps. H.263 was then defined to provide low bit rate video compression at rates lower than 64 kbps [39].

Other standards which are a part of the MPEG family include MPEG-1, MPEG-2, and MPEG-4. MPEG-1 is intended for video compression at rates around 1.5 Mbps, while MPEG-2 is intended for transmission channels in the range of 5-10 Mbps [15]. The former is intended for storing and retrieving moving pictures from databases while the latter is the standard for compressing digital television [6]. In addition, MPEG-2 provides a two-layered encoding scheme; the base layer when decoded alone provides a basic quality of service while decoding of the second layer uses residual information to enhance the base video quality [22]. MPEG-1 can achieve compression rates corresponding to the 64 kbps to 1.5 Mbps range, however substantial deterioration of the video quality results. MPEG-3 was initially intended to support High Definition TeleVision (HDTV) but was abandoned due to support for current MPEG versions supporting the encoding of HDTV [31]. MPEG-4 is the standard for multimedia applications and uses a low end bit rate of 5-64 kbps when a Common Intermediate Format (CIF) is used with a frame refresh rate of 15 frames/sec [7]. The very low bit rate encoding ability of MPEG-4 is appealing, the definition of standards is expected to be completed in late 1998 [13].

Typical applications that use the H.263 standard include video conferencing which usually consist of a *talking head*, such as the Miss America video sequence, and therefore contain a limited amount of motion. Due to the spatial and temporal redundancies found in these types of video sequences, compression is achieved using motion estimation and the Discrete Cosine Transform (DCT) which decrease the resulting bit rate considerably. For example, the Quarter Common Intermediate Format (QCIF) which has a luminance resolution of 144 lines by 176 pixels/line and two chrominance parts each having one quarter of this resolution, would require approximately 304128 bits per frame: 144 lines * 176

pixels/line * 1.5 (for the luminance and chrominance information). However, compressing the Miss America video clip, (with a Quantization Parameter (QP) of 1 to encode the most information) results in a mean bit rate of 46871 bits per frame; a compression gain of nearly 6.5.

Quantization is an irreversible operation that reduces the number of intensity levels found in the source image. The quantization parameter is used to adjust the amount of detail encoded in each frame and thus affects the resulting bit rate output by the video encoder. Increasing the QP reduces the amount of information and is similar to reducing the number of bits used to encode each pixel. If the QP is set too high, the information lost causes the images to become blurred [35]. Information that is lost when frames are compressed in this way cannot be recovered. Therefore, video compression using the H.263 standard, results in lossy compression, implying that the quality of the decoded video images is inferior when compared to the original video sequence. Decreasing the QP increases the amount of data encoded and improves the quality of the decoded image but results in larger compressed frame sizes.

The H.263 standard separates each frame according to the hierarchy shown in Figure 2.1. The uppermost layer, is the picture layer which consists of a header followed by a series of data units referred to as Group Of Block (GOB) data, followed by a trailer indicating the end of the picture. Each GOB data unit consists of a GOB Header followed by a series of MacroBlock (MB) data units. The MB data units are comprised of a MB header followed by a series of Block (B) data units. Each of the block data units are comprised of the transform coefficients followed by an End Of Block (EOB) sequence. The size of these structures is discussed below; since the video sequences we are using are stored in QCIF format, the discussion is also based on this format. QCIF has a resolution of 144 lines by 176 pixels. A block consists of 64 (8x8) pixels and a MB is comprised of 4 (2x2) blocks. Using QCIF, there are 99 (9x11) macroblocks in each frame, the numbering of the macroblocks is shown in Figure 2.2. A GOB in QCIF consists of a single row of 11 (176/16) macroblocks. Therefore, there are 9 (144/16) GOB data units in the picture layer when QCIF format is used.

The H.263 video compression algorithm can be separated into 4 distinct modules: motion estimation, transform coding, quantization of the coefficients and run length encoding. Intra-encoded frames do not use motion estimation since they exploit only the spatial redundancies present in the frame. However, prediction encoding references the previous encoded frame in order to exploit the temporal redundancies present. When prediction

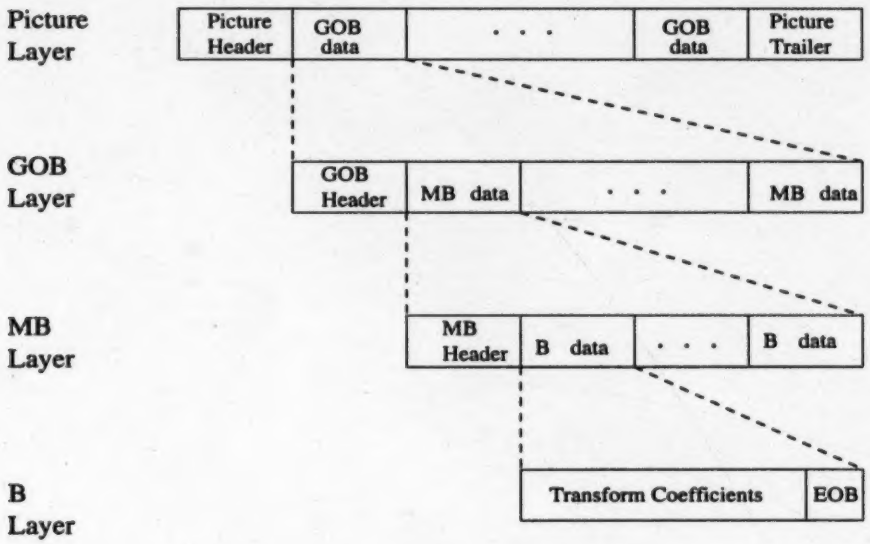

Figure 2.1: The H.263 hierarchical structure [35].

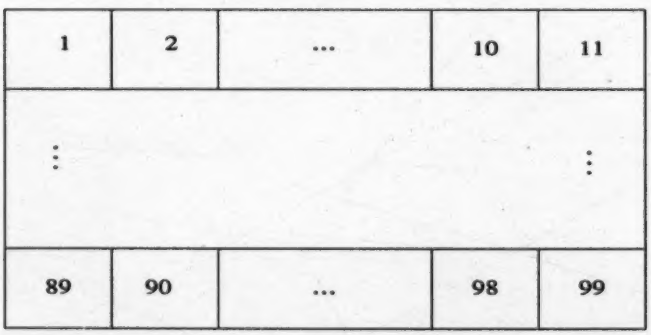

Figure 2.2: The numbering sequence of the 99 macroblocks found in an H,263 video frame when QCIF is used.

encoding is used, a search is performed using the surrounding macroblocks in the previous frame in order to find the closest match with the current macroblock being encoded. If the match is acceptable, the macroblock is encoded using the difference between the two macroblocks and the corresponding motion vector. Otherwise, the macroblock is intra-encoded. A match is thought to be close enough if the number of bits required to encode the difference is less than if intra-encoding is used. Bidirectional encoding is similar to predictive encoding except that the search for the closest macroblock is expanded to also include the P-frame currently being decoded. DCT coding is then performed on the coefficients of each block, followed by quantization of the coefficients, and finally, run length encoding to achieve further compression.

The allowable number of inter-encoded frames to follow an intra-encoded frame in the MPEG standard is defined by the Group Of Pictures (GOP) parameter which specifies the frequency that frames are intra-encoded [29]. If the GOP parameter is *N* then *N-1* inter-encoded frames would follow each intra-encoded frame, the value of N is usually between 12-15. Increasing the frequency that frames are intra-encoded reduces the propagation of errors, however, this results in more bits being generated. I-frames in the MPEG standard are particularly useful for video browsing [39] since a single frame can be decoded because it does not reference other frames.

2.2 CBR and VBR Video Encoding

In the following section we discuss the advantages and disadvantages of CBR and VBR video encoding. The results of the experiments run to evaluate the performance of the video encoder's CBR encoding option are discussed in Section 2.2.2. The results show that the CBR encoding option does not provide uniform frame sizes as intended, instead the burstiness of the bit stream produced increases.

2.2.1 A Comparison of CBR and VBR Encoding

Due to the spatial redundancies (data redundancies within a single frame) and temporal redundancies (associated with the degree of motion between successive frames) the compressed video stream created by the H.263 encoder may be highly variable. H.263 compression applied to a video clip with very little spatial detail and little or no motion will require far fewer bits per encoded frame, than frames consisting of large spatial detail and/or high motion such as high action scenes of a football game or panning of a video camera. Panning is the rotation of the video camera which results in global motion, whereas

motion associated with the movement of an object between adjacent frames is defined as local motion.

The high variability of spatial and temporal data to be encoded results in a highly variable number of bits associated with each frame. If the video encoder makes no attempt to modify the resulting bit rate, variable bit rate encoding results. Alternatively, the video encoder may dynamically modify internal parameters such as the amount of detail that is encoded in order to smooth the burstiness of the bit stream; this is constant bit rate encoding. Variable bit rate encoding is preferred over constant bit rate encoding since the picture quality remains constant [12]. However, although VBR video encoding is capable of providing constant picture quality, the large variability in the number of bits corresponding to each frame results in bursty traffic which is undesirable from a networking viewpoint.

Alternatively, when constant bit rate encoding is used, a sudden increase in the compressed bit rate would require modification of a parameter such as the quantization to be altered, causing less detail to be encoded in order to lower the corresponding bit rate leading to a degradation in the picture quality. Given these tradeoffs, from a video encoding perspective, variable bit rate encoding is preferred.

2.2.2 Evaluation of the H.263 Constant Bit Rate Option

In this section we show that the CBR option of the H.263 video encoder used in our experiments does not produce a CBR data stream. Section 4.3 of the H.263 standard [11] indicates that several parameters can be modified in order to alter the compressed video bit rates; one of these is the Quantization Parameter (QP). The encoder we are using for our experiments provides an *OFFLINE_RATE_CONTROL* option, which adjusts the QP in an attempt to smooth the variability of the frame sizes caused by the varying amounts of spatial and temporal detail present. This is the H.263 video encoder's CBR option.

When the CBR encoding option is invoked, the video encoder dynamically adjusts the Quantization Parameter (QP) which in turn modifies the level of detail encoded and the number of bits produced. Figure 2.3 shows a flowchart illustrating the information used to adjust the QP. Each time a frame is transmitted, a computation is performed to verify whether the QP should be modified. If the size of the last frame encoded exceeds the estimated size of remaining frames (based on the number of bits available to encode the remaining frames) by more than 15%, the QP is increased causing less information to be encoded and thereby decreasing the number of bits produced. Similarly, if the size of the last frame encoded is less than 85% of the frame size estimated, the QP is decreased to

increase the size of the subsequent frames in order to achieve the bit rate specified by the user.

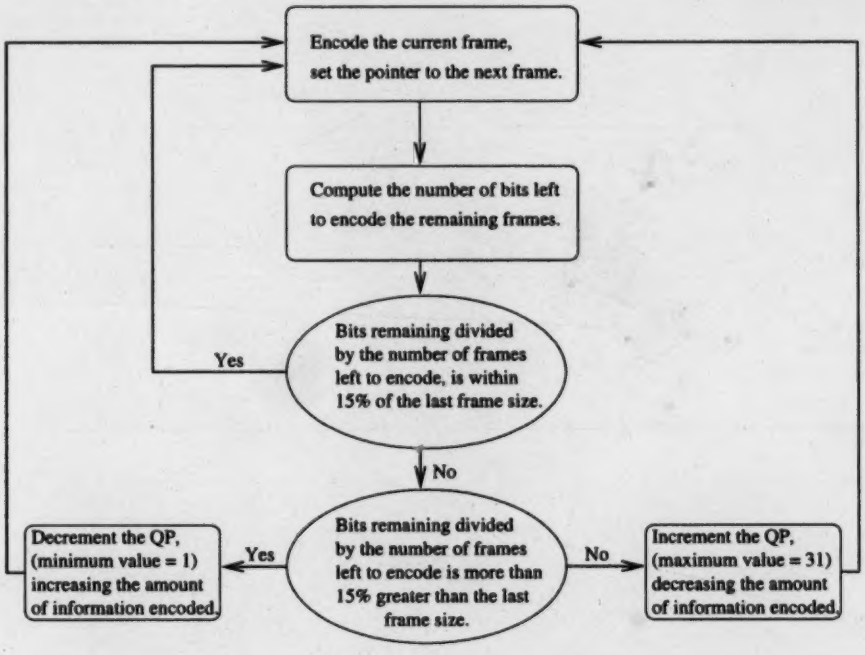
Encode the current frame, set the pointer to the next frame.
Compute the number of bits left to encode the remaining frames.
Bits remaining divided by the number of frames left to encode, is within 15% of the last frame size.
Yes
No
Bits remaining divided by the number of frames left to encode is more than 15% greater than the last frame size.
Yes
No
Decrement the QP, (minimum value = 1) increasing the amount of information encoded.
Increment the QP, (maximum value = 31) decreasing the amount of information encoded.

Figure 2.3: A flowchart illustrating the H.263 CBR encoding algorithm.

The Miss America video sequence has the lowest amounts of spatial and temporal information of the five video sequences used in our experiments. Therefore, this video sequence was chosen for this experiment, since video sequences with low amounts of temporal information should provide the best results when CBR encoding is used. We try to maintain a quality assessment value of at least 4, since this value corresponds to a subjective rating when there is a perceived degradation in the resulting video by the viewer, but it is not considered annoying. We found in our experiments that when VBR encoding is invoked on the Miss America video sequence with QP=7, the resulting quality is 4.286 and the corresponding bit rate is 44.45 kbps. Using this bit rate as a benchmark, we increase the bit rates by various percentages in order to investigate the bit rate stability when CBR encoding is used.

For consistency with the experiment using VBR encoding, we have initialized the QP to 7 when using the CBR encoding option. The experiments in this section intra-encode only the first frame and the remaining 149 frames are inter-encoded P-frames. This does

not adhere to the H.263 recommendation that macroblocks should be intra-encoded at least once every 132 times. However, the performance of the CBR option of the encoder would be further reduced if all the remaining frames were not inter-encoded because the intra-encoded frame sizes are much larger.

Several experiments are run using the CBR encoding option to evaluate if and at what point the frame sizes produced would have a variability at least resembling that when the VBR encoding is used. Therefore, the target bit rate is increased in each subsequent experiment and the resulting frame sizes and video quality are recorded. The experiments run and the target bit rates used in order to evaluate the performance of the CBR encoding option include: VBR (44.5 kbps), CBR (44.5 kbps), CBR (48.9kbps), CBR (51.1 kbps), CBR (80.0 kbps), CBR (489.0 kbps) and CBR (1378.1 kbps). These experiments correspond to bit rate increases of 0, 10, 15, 80, 1000, and 3000 percent over the corresponding bit rates when VBR encoding is used.

Table 2.1 summarizes the results when various target bit rates are used with the CBR encoding option. When the target bit rate is increased, the video encoder decreases the quantization parameter in order to utilize the available bit rate capacity, allowing more video information to be encoded. As shown in the table, increasing the target bit rate above the average bit rate that results when VBR encoding is used leads to more variability among the inter-encoded frame sizes. Comparing the variability of the frame sizes for the different bit rates is done using the *coefficient of variation* which is computed using Equation 2.1. The increased variability of the inter-encoded frames is due to the encoder modifying the QP while encoding is carried out.

Target Bit Rate	Increase Over	P-Frame Size			Video
(kbps)	VBR (%)	Mean (bits)	STD (bits)	Coef. Var.	Quality
VBR: 44.5	N/A	1400	370	0.07	4.286
CBR: 44.5	0	1410	490	0.12	4.160
CBR: 48.9	10	1552	530	0.12	4.269
CBR: 51.1	15	1630	590	0.13	4.290
CBR: 80.0	80	2588	1135	0.19	4.374
CBR: 489.0	1000	16467	15103	0.84	4.467
CBR: 1378.1	3000	44891	8495	0.04	4.589

Table 2.1: A comparison of the frames sizes when CBR encoding is used to compress the Miss America video sequence.

$$coefficient\ of\ variation = variance/mean^2 \tag{2.1}$$

The resulting video quality is lower using CBR encoding than when VBR encoding is used until the target bit rate is increased by 15% over the mean bit rate. The video quality is initially degraded due to some frames being encoded using larger QPs causing less detail to be encoded. However, when the target bit rate is increased by 15% enough information is being encoded that the video quality exceeds the result obtained using VBR encoding.

It is not until the target bit rate is set high enough that the quantization remains at 1 (thus encoding the greatest amounts of detail) that the P-frame size variability is less than when VBR encoding is used. This occurs at a target bit rate of 1,378.1 kbps. However, the actual bit rate is 1,340.5 kbps because the QP could not be decreased below 1 (the minimum), and therefore the maximum amount of information was sent. The cost to achieve this increase in video quality is a bit rate of nearly 31 times the bit rate of the VBR encoding.

The effects of increasing the available bit rate with respect to the resulting frames sizes are explained in the following figures. Figure 2.4 shows the frame sizes when VBR encoding is used with the QP=7. The first frame which is computed as an I-frame requires 13,640 bits and therefore appears as an outlier in the figure, while the remaining 149 P-frames have a mean of 1,400.2 bits.

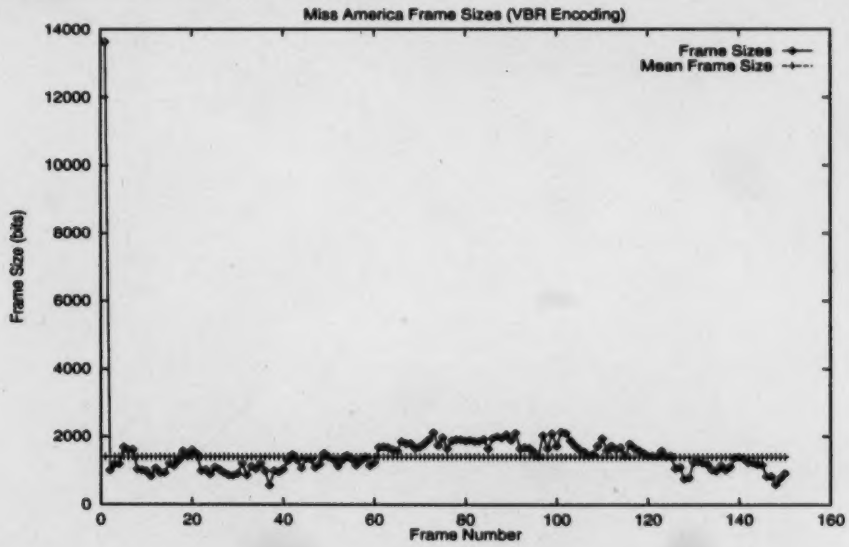

Figure 2.4: Miss America frame sizes using VBR encoding.

Figure 2.5 shows the frame sizes when CBR encoding is used with a target bit rate of 44.5 kbps (the average bit rate when VBR encoding is used). The first and last frames

appear as outliers in the figure because the first frame is intra-encoded while the QP is decreased for the last frame in order to utilize the remaining capacity of the target bit rate. Frames 64-70 and 71-86 are encoded using a QP of 8 and 9, respectively, as shown in Figure 2.6. Because the QP remains stable among these frames, their frame sizes are fairly consistent, while the remaining frame sizes are more variable due to the oscillations of the QP. Frame numbers 62-120 reflect a period when substantial amounts of temporal information are encoded causing the QP to be increased to the values greater than 7, the QP corresponding to when VBR encoding is performed. Increasing the QP is an attempt by the CBR encoding algorithm to reduce the larger frame sizes resulting from the larger amounts of motion present in the video sequence.

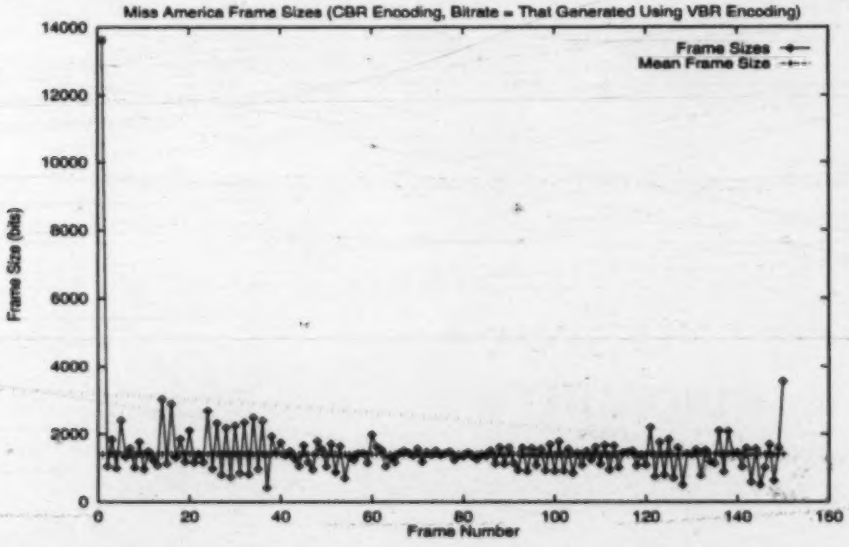

Figure 2.5: Miss America frame sizes when CBR encoding is used, (encoding bit rate = mean VBR encoding bit rate).

Figure 2.7 illustrates the frame sizes when CBR encoding is used after the bit rate is increased to 10 percent greater than when VBR encoding is used. The large variability in adjacent frames sizes particularly for frame numbers 23-37 are caused by the oscillation of the QP between the values of 5 and 6 as shown in Figure 2.8.

Figure 2.9 illustrates the frame sizes when CBR encoding is used after the bit rate is increased to 15 percent greater than when VBR encoding is used. As shown in the figure the frame sizes become much more variable compared to before. This is due to a larger number of frames being encoded while the QP oscillates between the values of 5 and 6 as

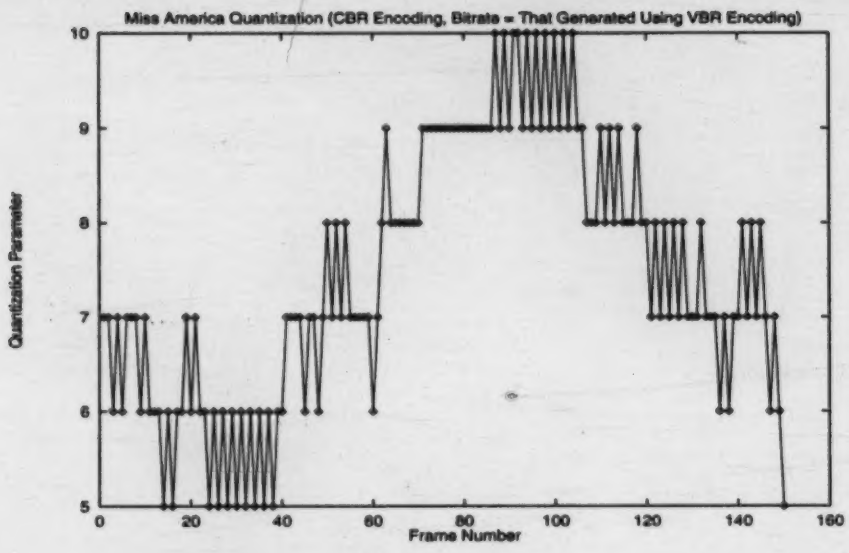

Figure 2.6: Miss America quantization parameters when CBR encoding is used, (encoding bit rate = mean VBR encoding bit rate)

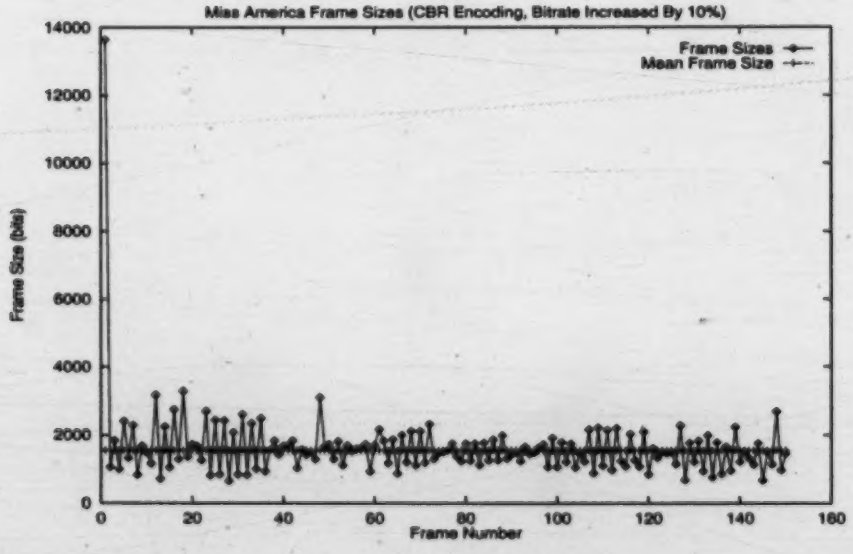

Figure 2.7: Miss America frame sizes when CBR encoding is used, (encoding bit rate = mean VBR encoding bit rate plus 10 percent)

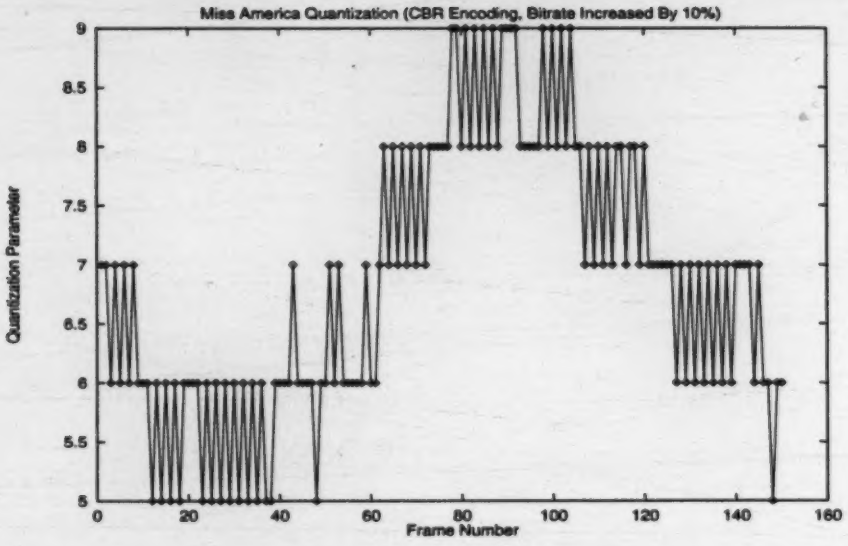

Figure 2.8: Miss America quantization parameters when CBR encoding is used, (encoding bit rate = mean VBR encoding bit rate plus 10 percent)

shown in Figure 2.10. For example, frame numbers 55-58 were encoded using a constant quantization of 6 when the target bit rate was 10% above that used with the VBR encoding. However, increasing the target bit rate an additional 5% has caused these adjacent frames to be encoded using QPs of 5 and 6 resulting in an increased variability of the resulting frame sizes.

Figures 2.11 and 2.13 show the frame sizes when the target bit rate has been increased by 80% and 1000%, respectively, over that of the bit rate when VBR encoding is used. The increased variability of frame numbers 29-39 in Figure 2.11 is due to the QP oscillating between the values of 3 and 4 as shown in Figure 2.12. When the target bit rate is increased by 1000% over that used with VBR encoding, the sizes of frame numbers 8-61 become highly variable since the QP now varies between 1 and 2 as shown in Figure 2.14.

As shown in Figure 2.15, once the target bit rate has been increased by nearly 31 times the rate produced by the VBR encoding scheme, the large variability among the P-frame sizes is reduced. Since the target bit rate is now large enough to permit the encoding of the frames using a QP of 1 as shown in Figure 2.16 the oscillations of the QPs are eliminated. Frame numbers 8 through 150 are encoded using a QP of 1 while the first two frames are encoded using a QP of 7 since the intra and inter QP variables are initialized to 7 for consistency with the previous experiments. The QP then decrements (the maximum QP

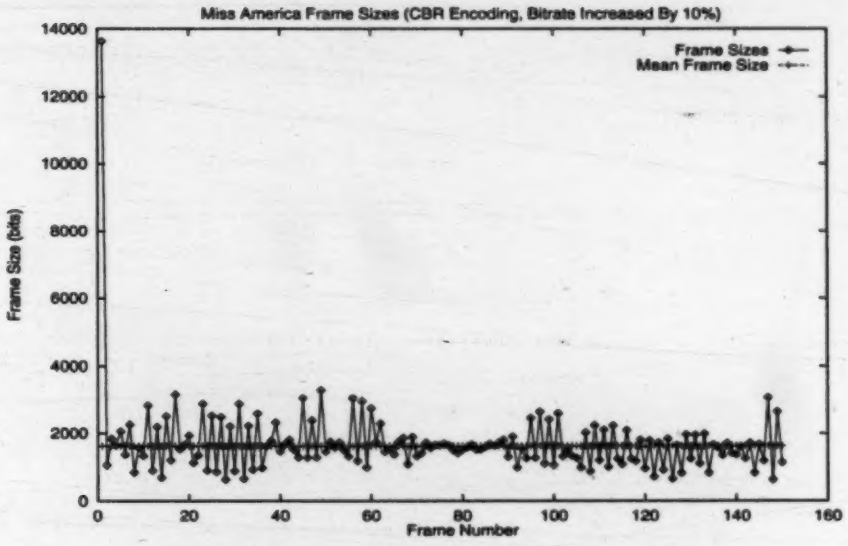

Figure 2.9: Miss America frame sizes when CBR encoding is used, (encoding bit rate = mean VBR encoding bit rate plus 15 percent)

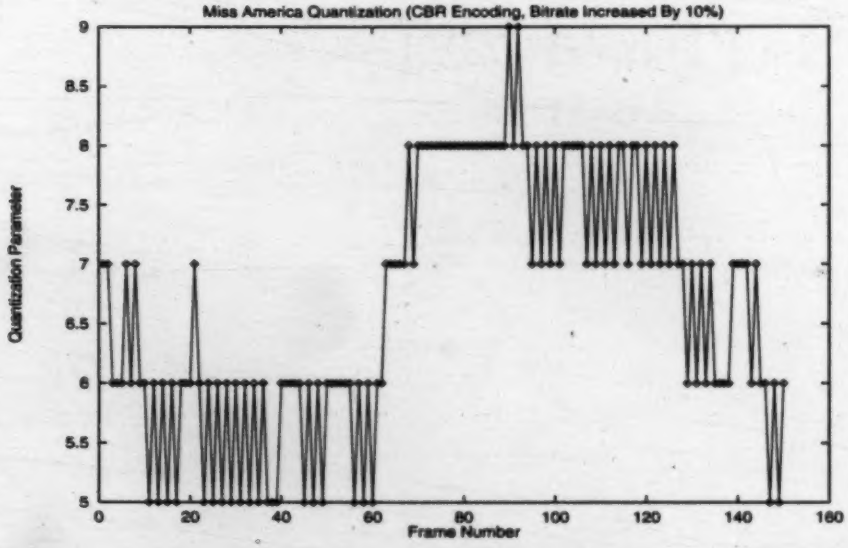

Figure 2.10: Miss America quantization parameters when CBR encoding is used, (encoding bit rate = mean VBR encoding bit rate plus 15 percent)

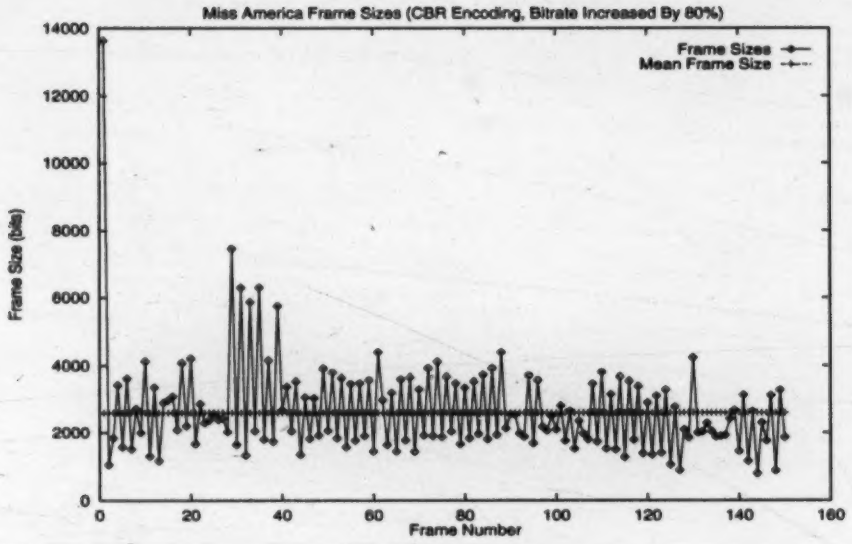

Figure 2.11: Miss America frame sizes when CBR encoding is used, (encoding bit rate = mean VBR encoding bit rate plus 80 percent)

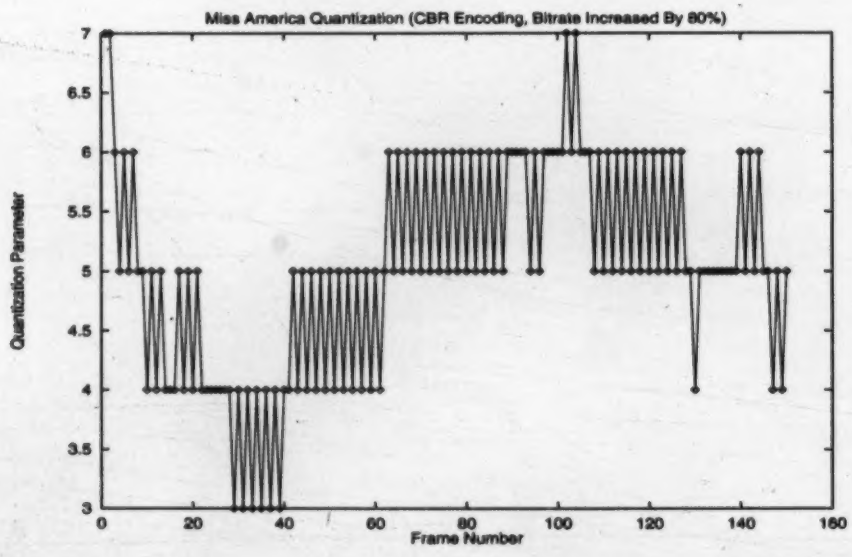

Figure 2.12: Miss America quantization parameters when CBR encoding is used, (encoding bit rate = mean VBR encoding bit rate plus 80 percent)

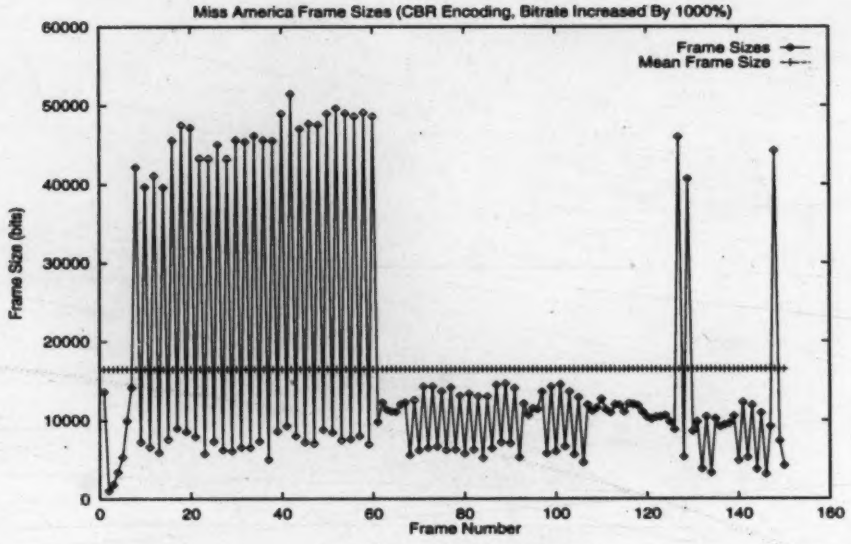

Figure 2.13: Miss America frame sizes when CBR encoding is used, (encoding bit rate = mean VBR encoding bit rate plus 1000 percent)

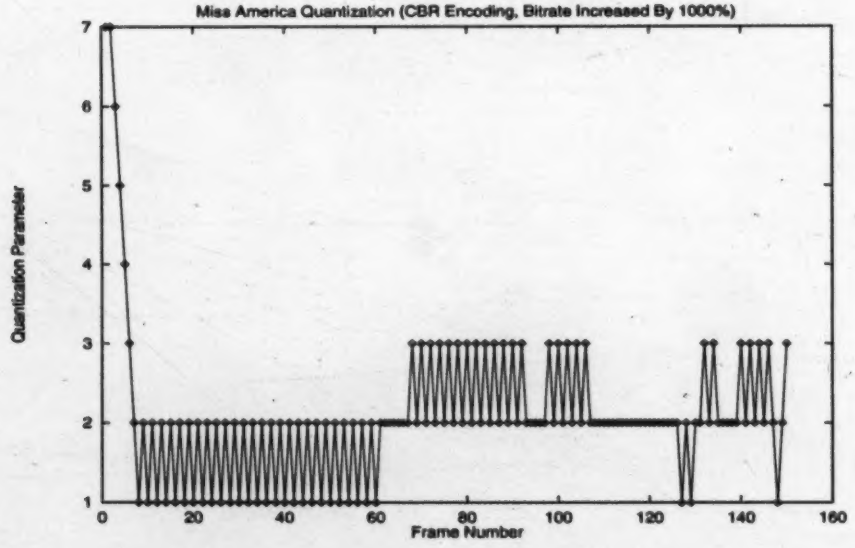

Figure 2.14: Miss America quantization parameters when CBR encoding is used, (encoding bit rate = mean VBR encoding bit rate plus 1000 percent)

change allowed per frame is 1) for each of the subsequent frames until the minimum of 1 is reached.

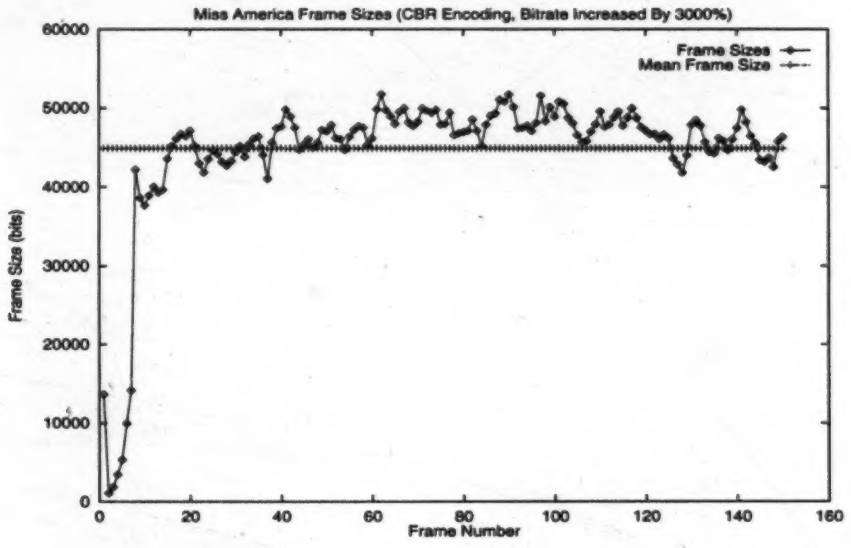

Figure 2.15: Miss America frame sizes when CBR encoding is used, (encoding bit rate = mean VBR encoding bit rate plus 3000 percent)

In Figure 2.15, the bit rate has increased to the point that the encoder is no longer able to increase the consumption of the available link by decreasing the QP since it is minimized to 1 throughout the video sequence. The variability in the frame sizes now is attributed to only the spatial and temporal data present. Since the QP value is 1, versus 7 when VBR encoding is used, the frame sizes are much larger. Similarly, small changes in the temporal or spatial detail present will result in larger frame size variability since the QP value is now 1 and the compression is not as significant.

In summary, the H.263 CBR encoding option does not produce uniform compressed frame sizes, instead it increases the burstiness of the bit stream produced. The burstiness of the bit stream produced continues to increase until the target bit rate is set high enough that the QP value remains constant at 1, thereby avoiding the oscillations encountered in the preceding experiments. If the amount of spatial and temporal information in the video sequence were consistent this method would perform much better than it does. However, even the Miss America video sequence which has the lowest amounts of spatial and temporal information present (see Table 5.1) results in very poor performance of this algorithm. Due to the problems illustrated here with the CBR option, we propose a video traffic smoothing

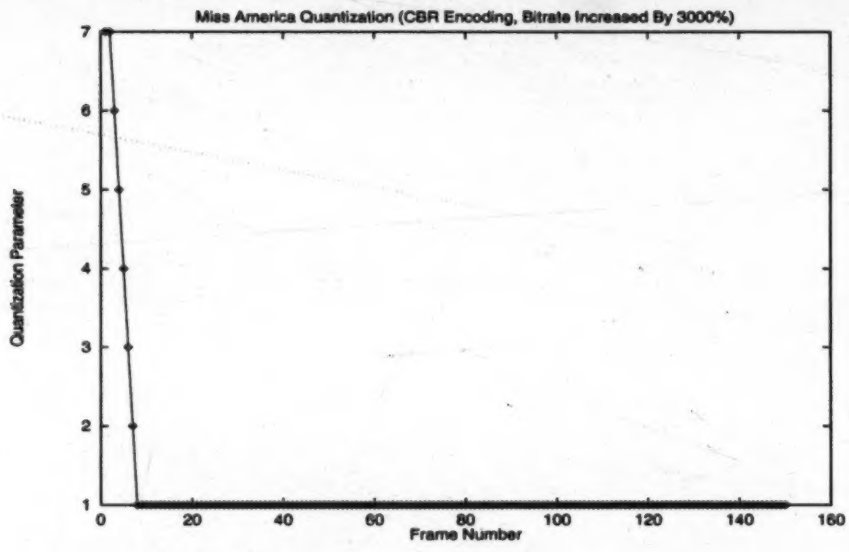

Figure 2.16: Miss America quantization parameters when CBR encoding is used, (encoding bit rate = mean VBR encoding bit rate plus 3000 percent)

scheme which will permit a VBR compressed video stream to be transmitted over a CBR network connection, while maintaining a high utilization and an acceptable video quality.

Chapter 3

Background Networking Information

In this chapter background networking information is provided which begins with an overview of Constant Bit Rate (CBR) and Variable Bit Rate (VBR) traffic from a networking point of view. In Section 3.2 we discuss previous work that has involved the transmission of compressed video over a variety of networks and outline the differences between our research and the work done previously. This is followed by a detailed discussion about the ATM networking protocol in Section 3.3.

3.1 A Networking Overview of CBR and VBR Traffic

From an image compression viewpoint, Variable Bit Rate (VBR) encoding is preferred over Constant Bit Rate (CBR) encoding (see Section 2.2.1). However, from a networking perspective, CBR traffic is easier to manage and guarantees that no information will be lost [18]. Consider for example a 1 Mbps portion of a network link where 10 users all wish to establish 100 kbps video connections. Using a Constant Bit Rate (CBR) traffic model, it is straightforward for a network provider to ensure each user will have the level of service they requested by allocating bandwidth equal to the user's specified peak bit rate for each CBR connection. If additional link capacity is requested, for example, by another user, and the aggregate bit rate of all the network users exceeds the capacity of the network, the user's request will be rejected.

Alternatively, when Variable Bit Rate (VBR) service is used, the bit rate requested by the user may be based on the average bit rate. Although the sum of the users' peak rates may exceed the capacity of the network allocated to VBR service, more users can utilize the network since the level of service is based on the sustained bit rate requested. However,

if the aggregate bit rate of the users at anytime exceeds the maximum network capacity, the data of one or more of the users would be discarded. Since VBR video traffic is self-similar, it tends to be very bursty and requires larger buffers at network nodes and decreases the network utilization. The bursty characteristic of self-similar traffic is scale invariant, therefore, aggregating traffic streams of this nature intensifies this bursty characteristic [9]. Traditional traffic models use a Poisson distribution to model the frame sizes. However, increasing the utilization of a link carrying self-similar traffic such as datagrams, can only be achieved if the buffer space is increased by orders of magnitude greater than those associated with traditional traffic models [3].

Because the CBR ATM service offers the user a dedicated connection, the user will desire a high link utilization should the network provider charge the user for the reserved network capacity rather than the capacity utilized. There are tradeoffs between a high link utilization and the probability of losses of transmitted data when a CBR network connection is used. In order to increase the link utilization, the CBR connection capacity is reduced and the user risks exceeding the peak bit rate capacity which would result in the loss of data and degradation of the video quality. Finally, when a CBR ATM network connection is used, the propagation delay and switching delays (on the order of a few microseconds) are bounded [2]; unlike VBR traffic over an ATM network which will experience varied network delays due to queuing at the internal nodes.

In order to decrease the variability of the frame sizes resulting from VBR encoding and the likelihood of exceeding the capacity of the CBR connection, *traffic smoothing* of the bit stream is required. Traffic smoothing can be achieved by implementing a buffer between the video encoder and the network source. Bits are then added to the buffer at a variable rate corresponding to the compressed frame sizes, while bits are removed at a constant rate determined by the capacity of the CBR connection. Since the aggregation of VBR video streams does not increase the link utilization, we use traffic smoothing to increase the utilization of the network connection thereby reducing the required CBR connection capacity.

The retrieval of the data from the buffer in the application layer by the ATM Adaptation Layer (AAL) is analogous to the *leaky bucket* algorithm which is used for traffic shaping (illustrated in Figure 3.1). The leaky bucket algorithm can be imagined as a bucket which has water flowing in at a non-constant rate, *e.g.* a smaller container is used to bail water into it, while the water flows out through a valve at the bottom at a regulated constant rate. In our experiments, the buffer corresponds to the bucket, the variable sized video

frames correspond to the smaller container used to fill the bucket while the ATM network cells remove data from the buffer at regular intervals, similar to the valve in the bottom of the bucket.

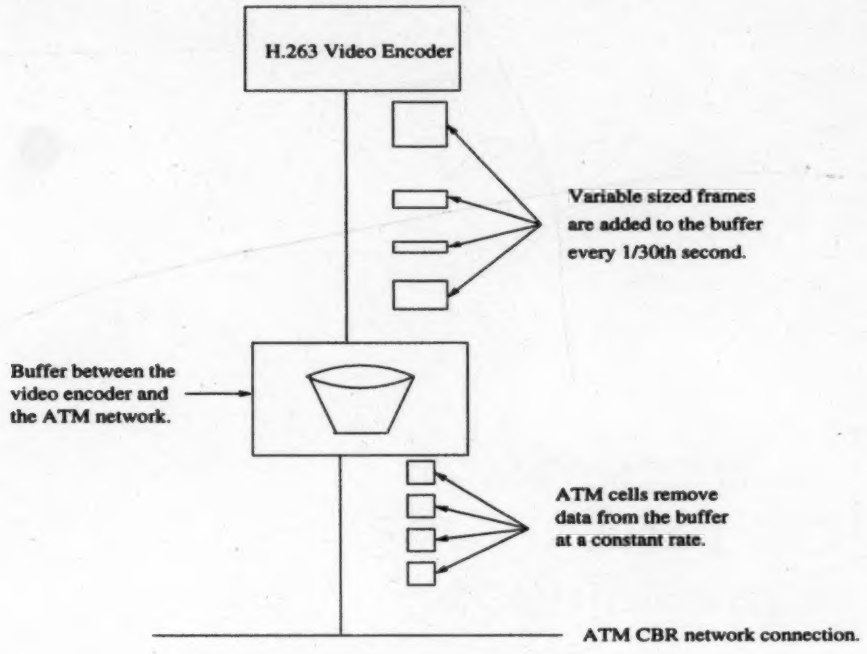

Figure 3.1: The leaky bucket algorithm [33].

3.2 Previous Work

In the literature extensive work has been presented on the performance of video conferencing over a variety of networks which include: the Internet, Ethernet, wireless links and ATM networks. In addition to the various networks, a considerable amount of work has also be done with respect to the various encoding algorithms which include H.261, H.263 and MPEG-2. After providing an overview of the work which has been done previously, we then describe how our work differs from the previous material.

The Internet is a collection of heterogeneous networks spanning the globe, providing shared network links instead of dedicated connections to users. Therefore, losses and delays may be expected should congestion occur within the network. Similarly, parameters such as the amount of loss and delay present are not known in advance since the network state is continually changing. The Resource reSerVation Protocol (RSVP) is intended to allow

applications to reserve the required quality of service in advance, although this service is not yet available [4].

Two transport layer (end-to-end communication) protocols are used in the Internet: User Datagram Protocol (UDP) and the Transmission Control Protocol (TCP). UDP is better suited for the transmission of real-time services such as videoconferencing applications than is TCP. The additional delays encountered with TCP (such as the retransmission of lost data) are unacceptable for real-time applications which includes video conferencing [35]. In addition, when modem connections are used the available bandwidth does not provide the capacity to permit the retransmission of lost data [39].

Ethernet LAN (Local Area Network) traffic is self-similar [16] and therefore multiplexing a number of these streams together increases the burstiness of the traffic rather than causing it to become smoother. Ethernet LAN can be described by the protocol used when a number of workstations are connected together on a single cable. Before a station transmits data, it listens to the cable to ensure it is not busy. If the cable is free, the station will transmit the data immediately, otherwise it will wait until the cable becomes free. When a station detects a collision of its data with that of another stations, it will terminate its transmission and wait a random amount of time before attempting to retransmit. This characteristic of Ethernet LAN traffic and the self-similarity of video traffic further affect the level of congestion on these networks. These characteristics adversely effect the network utilization and may lead to losses during peak periods as discussed in [16].

The lossy nature of wireless networks, (for example, during the *hand-off* of a host as it passes from one base station to the next), degrades the performance of applications using this transmission medium. These problems are particularly harmful to video applications since they utilize inter-encoded frames (frames that are encoded by referencing other frames in order to exploit temporal redundancies present and increase the amount of compression achieved). Previous work has investigated the retransmission of lost packets over a wireless network [1]. Liu and Zarki [21] discuss the ability to reduce the number of retransmissions required by using parity bits when transmitting H.263 over lossy wireless networks.

Extensive work has also been done experimenting with the transmission of compressed video using H.261 [8], [24], MPEG-1 [23], and MPEG-2 [10], [12], [22] over ATM networks. These works all share the central idea of transmitting VBR encoded video sequences over ATM networks. This work is taken a step further by Reininger *et al.* [28] who investigate bandwidth renegotiation during the transmission of MPEG-2 compressed videos. Many of these works evaluate the system performance in terms of buffering requirements at the

network source, the end-to-end delays and the assessment of the resulting video quality using the Signal-To-Noise Ratio (SNR). The signal-to-noise ratio measures the amount of useful information transmitted compared to the amount of noise present, lower SNR values indicate poorer signal quality [27].

Our work differs from previous work since we investigate the results when intra-encoding is done at the frame level, as well as when macroblocks rather than when entire frames are intra-encoded. In addition, instead of evaluating the degradation of the video quality due to lossy compression (and possible frame losses if buffer overflow is encountered) using the SNR, we use the video quality assessment described in [37] which provides a large correlation coefficient of 0.94 with the video quality perceived by the viewers as discussed in Section 5.2.1. Because the end-to-end delay in a CBR ATM network is nearly constant [2], we monitor the jitter caused by traffic smoothing at the network source.

3.3 Asynchronous Transfer Mode (ATM)

We have chosen to use an ATM network due to its popular and widespread use. Although the peripheral networks connecting most of the users are not ATM networks, the majority of the Internet backbones rely on the underlying ATM networks [26]. The original ATM standard indicated ATM should provide a data rate of 155.52 Mbps and an additional rate of 622.08 Mbps. These rates were selected since they are compatible with the Synchronous Optical NETworks (SONET) used by the TELephone COmpanies (Telcos). ATM technology chops data coming from the application layer above into 48 byte pieces and attaches a 5 byte header making up the 53 byte ATM cells which are then transmitted usually over fiber optic links. Because ATM does not have any specific physical layer characteristics, the cells are transported using SONET or some other transmission protocol.

Figure 3.2 illustrates the various fields found in the ATM cell header at the ATM layer. Figure 3.2 (a) corresponds to the User-Network Interface (UNI) used between the host and the ATM network; while Figure 3.2 (b) shows the cell header associated with the Network-Network Interface (NNI). The UNI is used for interaction between the customer and the ATM network, while the NNI is used for interaction between ATM switches within the network.

The General Flow Control (GFC) field consists of 4 bits and is specific only to the UNI. Although it was originally expected to provide flow control within the network, this functionality is not utilized [33]. The Virtual Path Identifier (VPI) consists of 8 bits if the UNI is used or 12 bits when the NNI is used. The Virtual Circuit Identifier (VCI) field

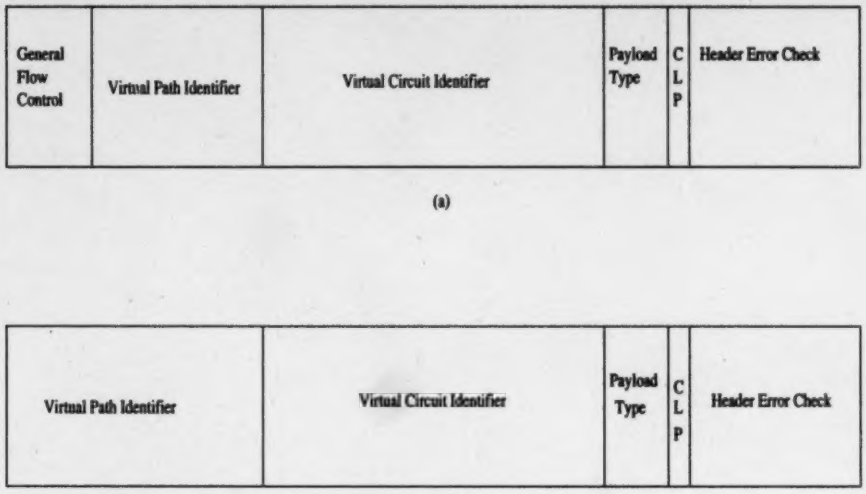

Figure 3.2: The ATM cell header formats [33].

consists of 16 bits, which specify the particular virtual path and virtual circuit to be used. Virtual paths and virtual circuits form a transmission hierarchy within an ATM network to simplify the network management. The transmission hierarchy is as follows: multiple VPs are multiplexed on a single physical link while multiple VCs are multiplexed onto a single VP. For example, a user running several different remote applications on the same host would utilize distinct virtual circuits for each application, but all of the user's connections to the remote host could utilize the same virtual path.

The last three fields of the ATM header consist of the payload type identifier, Cell Loss Priority (CLP) and the Header Error Check (HEC). The payload type field consists of three bits which can indicate whether the cell contains user data or maintenance information and congestion encountered in the network between the source and destination nodes. The CLP field consists of a single bit to indicate if the cell priority is high (1) or low (0). If congestion is encountered within the network, the network will discard the low priority cells during the congested period and attempt to forward the high priority cells. The last field, the HEC consists of a checksum over the entire cell header. This checksum is computed at each hop on the header alone in order to ensure cells are routed to the proper destination without the space and processing overhead which would be required to check the payload, which is substantially larger. Since 8 bits are reserved for the checksum on 40 bits, the algorithm is capable of correcting all single bit errors and approximately 90 percent of the errors caused

by multiple bit errors [33].

Figure 3.3 illustrates the layers and their locations in the ATM reference model. The physical layer (the lowest layer in the ATM reference model) consists of two sublayers, the Transmission Convergence (TC) sublayer and the Physical Medium Dependent (PMD) sublayer below it. The PMD is responsible for putting the bits on and off the fiber optic cable. The TC sublayer in turn is responsible for streaming the bits to the PMD sublayer. Because the role of the TC sublayer is to provide a uniform interface to the ATM layer above, it is also responsible for determining the cell boundaries when incoming bits are received from the PMD sublayer.

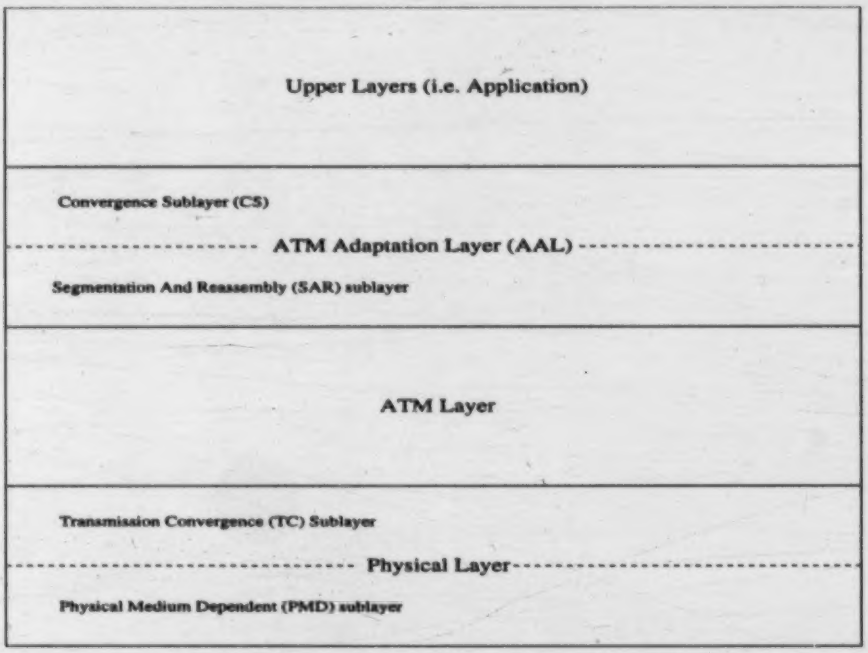

Figure 3.3: The ATM reference model [33].

The ATM layer provides a connection-oriented service between the source and destination using virtual circuits. Although services offered by virtual circuits are unidirectional, two virtual circuits can be established at call setup in order to provide full duplex communication. Communication offered by the ATM layer differs from other protocols (such as Internet Protocol) since it does not provide acknowledgments. This is mainly due to the following reasons: the probability of a bit error on a fiber link is very low, and the processing time is reduced. The probability of a bit error in a fiber optic network is about

10^{-9} [10]. Because ATM networks are designed for non-real-time and real-time applications, the additional delay associated with acknowledged traffic would not be acceptable for the real-time applications.

The ATM Adaptation Layer (AAL) provides the functionality of an interface between the applications using the ATM technology and the ATM network. The AAL consists of two layers, the Convergence Sublayer (CS) and the Segmentation And Reassembly (SAR) sublayer located below it as shown in Figure 3.3. The CS provides an interface to the application and is responsible for breaking the data stream provided by the application into 48 byte pieces (or smaller) corresponding to the cell payload which is protocol dependent. The SAR sublayer then passes these payloads to the ATM layer for transmission. At the destination when the SAR sublayer receives cells, it is responsible for reassembling them into messages which are passed onto the CS.

There are various classifications of AALs which include: AAL 1, AAL 2, AAL 3/4 and AAL 5. The AAL 1 protocol is intended for real-time, connection-oriented, CBR traffic which includes uncompressed video and audio applications. Since this service is intended to minimize the delay, jitter and overhead, error detecting protocols are not used. This protocol attaches a 1 byte header at the SAR sublayer to detect lost or missing cells, thus reducing the cell payload to 47 bytes. The optional *P cells* use a pointer field consisting of 1 byte, to specify the offset of the next message boundary. These are not required in our application since the message boundaries are recognized by the Picture Start Code (PSC) defined in the H.263 standard. Reports of missing or lost cells are provided to the application which is then responsible for initiating its own action to recover them if desired.

The AAL 2 protocol differs from the previous protocol in that it is intended for variable bit rate applications such as compressed video. After compression, the resulting bit rates may vary significantly depending on the amounts of spatial and temporal information present. In addition, video compression algorithms often reference other frames in order to exploit redundancies associated with motion between adjacent frames. Therefore this protocol adds a 1 byte header and 2 byte trailer to the cell payload (of 45 bytes) in order to provide information which includes the sequence numbers of each cell and a checksum over the entire cell. Due to a number of problems (*e.g.*, the field sizes are not defined in the standard) associated with the AAL 2, this protocol is not used [33].

The AAL 3/4 protocol is defined for non-real-time variable bit rate applications requiring connection or connectionless services. This protocol unlike the previous two offers multiplexing capabilities. Since network vendors often charge for the time a connection is

established, this functionality allows the user to have multiple activities such as remote logins or file transfers on a VC between a common host/destination pair. The level of service among the user's applications would then be equal since virtual circuits can provide only one level of service at a time. A Multiplexing IDentification (MID) field comprised of 10 bits is part of the cell payload header, and is used to separate the cells pertaining to different user sessions at the destination.

The AAL 5 protocol, formerly known as the Simple Efficient Adaptation Layer (SEAL) was designed by the computer industry unlike those previously mentioned which were developed mainly by the telecommunications industry [33]. A significant advantage of the AAL 5 over the AAL 3/4 protocol is with respect to efficiency. While the AAL 3/4 protocol adds 4 bytes to every cell and 8 bytes to every message, the AAL 5 protocol adds 8 bytes per message but does not add any overhead to the cell payload. Therefore, applications using large messages such as datagrams which can be up to 65535 bytes in length, the overhead is significantly reduced. In addition, the protocol-processing overhead is reduced since the AAL 5 protocol is optimized to provide connection-oriented services since the unnecessary services pertaining to the connectionless services of AAL 3/4 are discarded [32].

Although, both AAL 1 and AAL 5 have been proposed for the transport of MPEG compressed video, AAL 1 is superior since it is able to resynchronize the data's clock with the timing at the receiver, providing the near constant delay that the MPEG standard expects [10]. Modified versions of AAL 1 have been proposed where Forward Error Correction (FEC) is used in order to provide error robustness without the need for retransmission. The disadvantage of FEC is the overhead of the additional bits required in order to correct possible bit errors. However, the disadvantage of this lost network capacity may be outweighed in real-time applications where delays associated with the retransmission of data are not acceptable.

Since the AAL 5 is unable to minimize cell jitter [10] we propose to use the AAL 1 protocol. This is intended for real-time, connection-oriented applications such as video conferencing. Although AAL 1 is intended for CBR traffic, it will satisfy our requirements since our simulation is intended to smooth the compressed VBR H.263 video stream for transmission over a CBR network link.

Chapter 4

Design of our Video Encoder Algorithm

In this chapter we discuss two types of H.263 video encoders that we have designed: a black box video encoder and a dynamic video encoder. It is shown in Section 2.2.2 that the CBR encoding option of the H.263 encoder produces greater variability among the frame sizes than when VBR encoding is used. Therefore we provide a solution where the VBR encoded video can be transmitted over a CBR ATM network connection using a buffer to smooth the video traffic in order to increase the utilization of the CBR connection. The removal of data from the buffer at a constant rate by the CBR ATM connection smoothes the VBR video traffic; this traffic shaping uses the leaky bucket algorithm.

Consideration must be given to the implications which arise should buffer overflow be encountered. That is, knowing that inter-encoded frames reference the previously encoded frame, is it reasonable that the video encoder has the ability to monitor the buffer conditions? If not, the H.263 video encoder is treated as a *black box*; otherwise the encoder is considered to be *dynamic*. Section 4.1 provides an overview of our black box video encoding algorithm and a description of the implementation, followed by Section 4.2 which discusses our dynamic encoder algorithm and its implementation.

4.1 A Black Box Video Encoder

This section provides an overview of our black box video encoding algorithm, followed by a description of the required modifications to the H.263 video encoder source code for the algorithm's implementation.

4.1.1 An Overview of the Black Box Encoding Algorithm

Treating the H.263 video encoder as a *black box* is a reasonable assumption, because it

may be a piece of hardware simply intended to be plugged into a system and therefore would not provide the functionality required to allow user modifications. Figure 4.1 illustrates the main parts of this system which include: the video encoder, a buffer (that complete frames are added to) used for smoothing the compressed VBR traffic, and an ATM network which is used to transmit the data to the decoder. As shown in this figure, the video encoder is treated as a black box, and therefore, it is unable to monitor the buffer utilization in order to know if a frame is added to the buffer or discarded due to buffer overflow.

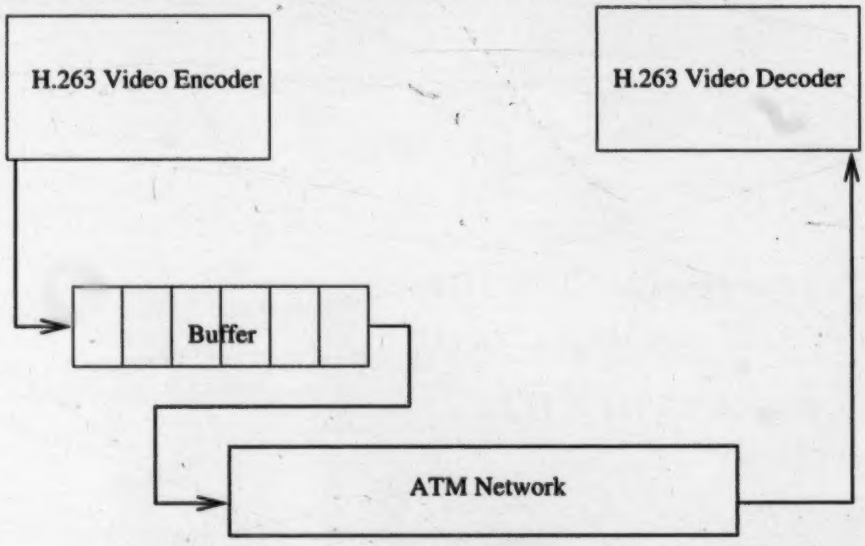

Figure 4.1: The H.263 video encoder as a black box.

Consider, for example, that frame *n-1* has just been encoded and added to the buffer as shown in Figure 4.2 (which illustrates the position of a ball in three consecutive frames). The arrows in the figure show the previous frame referenced during the encoding and decoding of the inter-encoded frames when frame *n* is discarded. Frame *n-1* is then decoded at the video encoder creating a reconstruction of frame *n-1*. We refer to the decoded frame as the frame's *reconstruction* since it is not identical to the original frame; some losses may have occurred due to the lossy H.263 compression algorithm. Frame *n* is then encoded by referencing the reconstruction of the previous frame, however, due to buffer overflow, the bits pertaining to the compressed version of frame *n* are discarded at the network source. Because the video encoder has no knowledge of the buffer utilization it is unaware that frame *n* has been discarded and therefore references the reconstruction of frame *n* when encoding frame *n+1*. Since frame *n* was discarded at the network source due to buffer overflow, using

the reconstruction of the last frame received (frame *n-1*), frame *n+1* is incorrectly decoded because it was encoded using the reconstruction of frame *n*. Further degradation of the video quality occurs since the errors are allowed to propagate until they are corrected using intra-encoding. In accordance with the H.263 recommendation, a macroblock could be inter-encoded up to 132 times before intra-encoding is used and these errors are corrected.

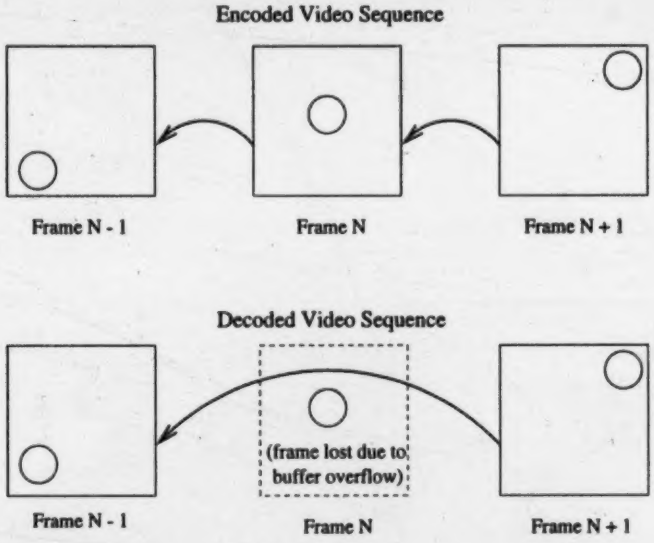

Figure 4.2: Encoding scheme when using a black box video encoder.

4.1.2 Implementation of the Black Box Video Encoder

We compute the video quality at the encoder by comparing the original frame quality and the reconstructed frame quality. This method is used since the H.263 encoder decompresses the frame in order to compute the SNR of the decoded video quality. Additional structures were implemented at the encoder for the purpose of assessing the degradation of the video quality experienced when frames are lost due to buffer overflow.

When a black box video encoder is used, inter-encoded frames are encoded referencing the reconstruction of the previous frame encoded. This reconstruction does not reflect any changes when a frame is discarded, because a black box encoder is unaware of the buffer conditions. Computing the video quality at the encoder requires referencing the same frame that would be referenced when a frame is decompressed at the decoder. This takes into account lost frames and requires keeping a copy of the reconstruction of the last frame successfully added to the buffer in order to decode the frame just encoded. The

reconstruction of the last frame successfully added to the buffer is analogous to the decoded version of the last frame successfully received at the decoder.

4.2 A Dynamic Video Encoder

Section 4.2.1 provides an overview of the dynamic video encoding algorithm. A description of the source code modifications required to implement the dynamic video encoder is provided in Section 4.2.2.

4.2.1 An Overview of the Dynamic Encoding Algorithm

The H.263 video encoder is *dynamic* if it has the ability to monitor the utilization of the buffer used to smooth the VBR data stream then the errors discussed in the previous section can be eliminated. As shown in Figure 4.3 a *feedback loop* is required allowing the encoder to monitor the buffer utilization. With knowledge of the buffer utilization, the application is aware when a frame will be lost and can encode the video sequence accordingly. That is, the encoder is capable of encoding the frames so that inter-encoded frames reference the same frame used during encoding and decoding regardless of whether a frame is lost due to buffer overflow.

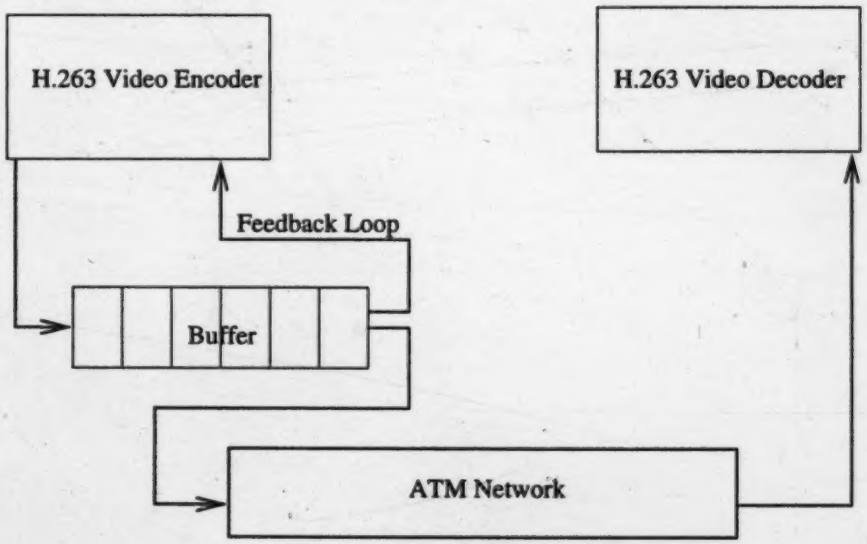

Figure 4.3: A dynamic H.263 video encoder.

Figure 4.4 shows the encoding scheme when a dynamic video encoder is used and frame

n is discarded. The arrows in the figure illustrate the previous frame referenced during the encoding and decoding of the inter-encoded frames when a frame is discarded. As shown in the figure, when frame *n* is discarded due to buffer overflow, frame *n+1* is encoded referencing the reconstruction of frame *n-1*. Although the information pertaining to frame *n* cannot be recovered since the frame is lost, frame *n+1* will be decoded correctly since it will reference frame *n-1*.

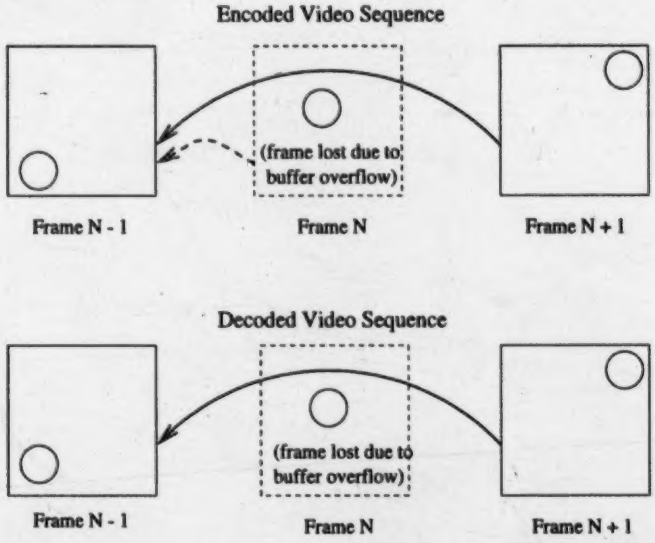

Figure 4.4: Encoding scheme when using a dynamic video encoder.

4.2.2 Implementation of the Dynamic Video Encoder

The video quality of the decoded video sequence is computed at the encoder as explained in Section 4.1.2. Implementing the dynamic video encoder required modifications to the video encoder source code. The dynamic H.263 video encoder we have implemented keeps a copy of the reconstruction of the last frame successfully added to the buffer, in order to decide what to reference when encoding the next frame.

After a frame is encoded, the buffer space is checked to verify whether the size of the current frame exceeds the current buffer capacity. If the current frame size exceeds the buffer capacity, the frame is discarded and a flag is set. Using this flag, the dynamic encoder does not update the pointer to the reconstruction of the last frame transmitted until the next frame is successfully added to the buffer. For example, if an I-frame is added to the buffer causing successive P-frames be discarded, the update of the pointer to the reconstruction of

the last frame transmitted will be delayed. I-frames are required to prevent the propagation of errors and are therefore never discarded as discussed in Section 6.2.2.

The dynamic video encoder avoids frames being incorrectly decoded, which happens when frames are lost and the black box encoder is used. The dynamic video encoder requires keeping track of a pointer to the reconstruction of the last frame transmitted and the current buffer capacity. The main disadvantage of this algorithm is that it requires modification to the source code; if the user does not have access to the source code, this algorithm cannot be utilized.

Chapter 5

Overview of Our Simulation Model and Assessment Criteria

This section provides an overview of our simulation model and the assessment criteria used in this thesis. The following section provides an overview of our simulation design. Section 5.2 discusses the assessment criteria used to evaluate the video quality and the network performance. Section 5.3 describes the five video sequences available for testing and outlines the two sequences chosen for our experiments.

5.1 Simulation Design

This section provides an overview of our simulation design. In the following section we provide a brief description of the system used in our experiments. Section 5.1.2 outlines the simulation events followed by an overview of the system's implementation in Section 5.1.3. In Section 5.1.4 we outline the simplifications we have made and the methods used to validate our system are discussed in Section 5.1.5.

5.1.1 Simulation Model

This section provides a description of the simulation model which is shown in Figure 5.1 (discussed briefly in Section 1.2). The H.263 video encoder is used to compress video sequences stored in files using the Quarter Common Intermediate Format (QCIF). The information for each frame in the file begins with the Y-component (luminance information) followed by the U and V components (the color information). In our experiments we use five standard video sequences, each consisting of 150 frames. Because the frame refresh rate used in our experiments is 30 frames/sec, compressed frames are added to the buffer every 1/30th of a second. A simulated buffer is used to smooth the VBR data stream. A simulated CBR ATM network connection then removes the data from the buffer at a constant rate

determined by the connection capacity. The traffic shaping in our system uses the leaky bucket algorithm since data is added to the buffer at a variable rate and removed at a constant rate. Because the video encoder does not run in real-time, the compressed video sequence is written to a file. This file is then input to the video decoder which displays the video sequence in real-time on a computer monitor.

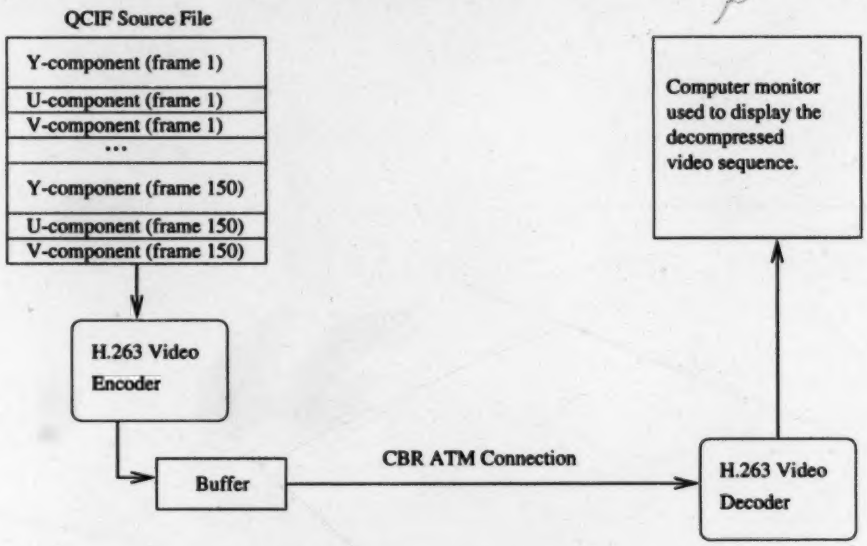

Figure 5.1: Overview of the system components used in our experiments.

5.1.2 Simulation Events

Our experiments use an H.263 video encoder implemented in software to compress video frames which are then transmitted across a simulated, CBR ATM connection to the video decoder. The simulation events are at the ATM cell level, rather than either the frame level or the bit level. Events at the cell level allow us to monitor the impact of the variability of the compressed frame sizes which are determined by the amounts of spatial and temporal detail present. This trace driven simulator adds a compressed frame to the buffer every 1/30th of a second (in simulation time) and removes a cell containing 47 bytes from the buffer at intervals of Δt. The inter-cell departure time, Δt, is dependent on the corresponding bit rate, since video clips with a higher bit rate will transmit cells more rapidly and thus have a lower Δt value than lower bit rate clips. Since the H.263 encoder does not run in real-time, the amount of processing time required to encode the video frames is not known. Therefore, frames are added to the buffer as complete frames rather than as a data stream of known

duration. Whenever a frame is added to the buffer or the payload of a cell removed, the buffer occupancy is updated.

In order to reduce roundoff errors, time is measured in microseconds. A simple calculation outlining the event granularity is given below. If the mean bit rate of a video is 80 kbps, cells would be transmitted as follows: 80 kbps $\Rightarrow$ 10000 bytes/sec $\Rightarrow$ 212.77 cells/sec $\Rightarrow$ cell inter-departure times of 4700 microseconds. Therefore, every 4700 microseconds a cell would traverse the network, regardless of whether it was filled to capacity; cells not filled to capacity adversely affect the network utilization.

5.1.3 Implementation

This section gives an overview of the implementation required in this thesis to transmit a compressed H.263 video stream over a simulated ATM connection. Our ATM network simulator and the H.263 video encoder/decoder are written in the *C* programming language. Important points pertaining to our implementation include a description of how the VBR data stream produced by the video encoder is smoothed, issues pertaining to frame losses, how the jitter is computed, and finally, computing the video quality at the video encoder.

The H.263 video encoder creates a VBR bit stream which is smoothed using a buffer that has data removed at a constant rate (the leaky bucket traffic shaping algorithm) by a simulated CBR ATM connection. Should buffer overflow be encountered, *e.g.*, when a very large I-frame is added to the buffer, frames are discarded based on the size of the frame; this is discussed in detail in Section 6.2.1.

There is no communication required between the encoder and decoder to provide notification of frame losses because the frames are displayed at the decoder in the order they are received. For example, when a very large intermediate I-frame is transmitted over a low capacity CBR connection, the delay associated with the transmission of this frame will result in a frozen scene at the receiver until the I-frame is received and the frame display is updated. Although this degradation is not captured by the video quality assessment algorithm the corresponding jitter computed will take this into account.

When low capacity network connections are used, the first frame that is intra-encoded delays the initial display of the video conference at the decoder. However, since it does not affect the inter-frame display time, it is not included in the jitter computation. For consistency, the *mean bit rate* discussed in the following experiments does not include the first I-frame in the computation. Therefore, the mean bit rate is the average bit rate computed using frame numbers 2-150 inclusive.

The video quality assessment algorithm computes the result using the original frame before compression, and a copy of the frame which has been compressed and then decompressed. Because the H.263 video encoder we are using decompresses each frame in order to compute the Signal to Noise Ratio (SNR), we also use this copy to compute the video quality at the video encoder. This is equivalent to computing the video quality at the decoder because we assume no errors are introduced in the fiber optic network, and no cell loss (due to congestion) is experienced within the CBR network connection.

5.1.4 Simplifications

Since we are simulating the ability of our proposed algorithm to smooth the VBR compressed video stream for transmission over the CBR connection-oriented service, our simulation uses a point-to-point connection. When an ATM CBR quality of service is granted, provided that the CBR capacity is not exceeded the network guarantees cell delivery without losses. Therefore we model the network as a single connection link of a specified capacity that is not affected by other network traffic.

The delay discussed in this thesis ignores the end-to-end delay which includes the propagation and switching delays encountered within the ATM network. Although the propagation delays may be as high as 100 msec, corresponding to a global propagation delay, they are dependent upon the physical distance between the two parties in the video conference, and therefore remain constant throughout the duration of the video conference session. The circuit switching delays are bounded, in the order of a few microseconds, because we are using a CBR quality of service [2]. The delay we measure corresponds to jitter, which is caused by the smoothing of the VBR traffic at the network source.

Although the H.263 encoder (Version 2.0 developed by Telenor R&D, Norway) that we are using does not run in real-time, it does allow us to study the network requirements and performance when various video sequences are compressed and then transmitted over a simulated ATM network. Since the encoder does not run in real-time, frames are added to the buffer as complete frames rather than as a data stream. The AAL 1 protocol then retrieves the bits at the cell level from this buffer which are transmitted over the ATM connection at a constant rate.

In order for users to reserve CBR network connections, they require some knowledge of the capacity required for their application. Computing the required CBR connection capacity in advance for applications such as video on demand providing coverage of a football game, which has periods of high and low action can be difficult. However, video conferencing

applications usually have much lower amounts of motion and the amount of spatial detail is much more uniform. Therefore the bit rate generated when the encoder is run for a short duration off-line may accurately determine the required capacity of the CBR connection.

5.1.5 Validation

Validation of our experiments is based on information found in current research papers and the mathematical analysis used to verify our experimental results. Considerable reading of current research papers (outlined in Chapter 1 and Section 3.2) related to the following topics have been used to formulate our experiments which include: transmission of compressed video over ATM networks, video quality assessment, VBR traffic smoothing for transmission over CBR connections.

The video sequences used in our experiments remove randomness from the input. Therefore, our experimental setup is deterministic rather than probabilistic and confidence intervals are not used. Our experiments are trace driven by the compressed frame sizes generated by the H.263 video encoder which encodes each frame on 1/30 second intervals. In addition, we do not require any random variables to specify cell inter-departure times since this is determined by the CBR connection capacity.

The video sequences used for video encoder input are not easily generated, therefore we use five standard sequences available, each one consisting of 150 frames. For brevity, the majority of our experiments utilize the two video sequences having the highest and lowest amounts of spatial and temporal information present, while our experiments in Section 7.3 use all five of the video sequences in order to substantiate our conclusions.

5.2 Assessment Criteria

This section describes the assessment criteria used in this thesis. The following section discusses the criteria used to evaluate the video quality in our experiments. Section 5.2.2 outlines the criteria used to evaluate the network performance.

5.2.1 Video Quality Assessment Criteria

In order to evaluate the performance of our experiments it is important to evaluate the resulting video quality using a method which provides results that closely correspond to the quality perceived by the user. In this section we discuss the algorithms used to evaluate the video quality in our experiments.

Video Quality Assessment Algorithm

Much of the research currently done in the area of video compression assesses the video quality using objective measures such as the Signal to Noise Ratio (SNR). However, we desire a method to assess the video quality which consistently provides results that correlate closely with the quality perceived by users. That is, an objective assessment which will provide results similar to those obtained using subjective testing. This method should have the advantages of objective testing such as quick and easy implementation, and the advantages of subjective testing in order to provide results similar to the ratings provided by users.

We use the video quality assessment proposed by Webster *et al.* [37] which provides an objective method of assessing the quality of a video clip, and provides results similar to those obtained by a panel of viewers. A panel of 48 viewers, whose names were listed in the Boulder, Colorado U.S. Department of Commerce Laboratories phone book were asked to rate the video clips with a rating score in the range of 1-5. The ratings represented: (5) Imperceptible, (4) Perceptible but not Annoying, (3) Slightly Annoying, (2) Annoying and (1) Very Annoying. The viewers were asked to view 38 or 40 video clips during each of four sessions over the period of one week. Each video clip lasted 30 seconds, consisting of 9 seconds of the original video, 3 seconds of grey display, 9 seconds of the degraded video followed by a 9 second rating period. The video clips ranged from still scenes to full motion entertainment video, and consisted of 36 scenes. 132 of a possible 972 video clips were used in testing, these were chosen both deterministically and randomly. Degradation of the original video streams was due to noisy Radio Frequency (RF) transmission and the bit errors encountered in the video systems used.

Figure 5.2 shows the process which was used in order to obtain the video quality assessment algorithm. Processing, such as first order differencing and Sobel filtering, were used to calculate features of the original and degraded video streams as part of the objective testing step. A Sobel filter is an edge enhancement filter used to measure the lost or gained edge energy which may appear as false or blurred edges in the degraded video clip. Single scalar values (parameters) of these features were then produced by collapsing them over time using operations such as the STandard Deviation (STD) and Root Mean Square (RMS). The parameters obtained formed the objective measurements which were then used in the statistical analysis step (shown in Figure 5.2) to obtain the final parameters. Using statistical analysis an algorithm was obtained to provide video quality ratings closely matching those of human perception. Video quality ratings using this method obtained a

correlation coefficient of 0.94 with the subjective ratings.

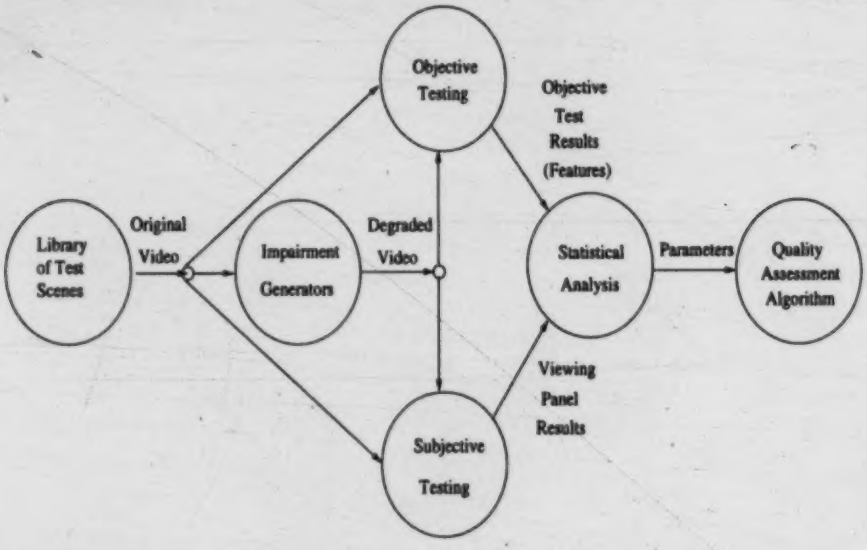

Figure 5.2: Development process for the video quality assessment algorithm [37].

Distortions perceived in degraded video streams can be divided into two main sources, spatial distortion and temporal distortion. Spatial distortions relate to false or blurred edges which may be caused by the amount of compression. For example, increasing the quantization parameter in the H.263 encoder will decrease the corresponding bit rate required to encode the video sequence, but will result in a reduction of the spatial detail present in the decoded video stream.

Temporal distortion relates to the degree of motion which appears to have been lost in the degraded video. Low amounts of temporal distortion may result in an appearance of uneven motion of objects in adjacent frames, but may be as extreme as a *frozen* display due to inadequate frame update rates. Although the spatial and temporal information perceived by the human vision system cannot be completely decoupled, the algorithm discussed here selects spatial and temporal features which correlate closely with the spatial and temporal information perceived by human vision systems.

The spatial and temporal distortions measured were obtained from the luminance portion of the signal. Although a variety of color distortion measures were analyzed, they were found to be insignificant in comparison to the spatial and temporal distortions, and therefore were not selected in the optimization algorithms used.

The algorithm proposed by Webster *et al.* derives three parameters which are then used

in a linear combination to provide a rating of the video quality on a scale of 1-5. The first parameter m_1 is a measure of the spatial distortion in the degraded video; while parameters m_2, and m_3 identify the degree of temporal distortion present.

The first parameter m_1, is obtained using the equation:

$$m_1 = RMS_{time}(5.81[(SI[O_n] - SI[D_n])/SI[O_n]]) \tag{5.1}$$

where *Spatial Information (SI)* is the standard deviation of the Sobel operator applied to the pixels in frame n of the Original (O) and Degraded (D) video sequences. Video quality degradation caused by spatial blurring will result in a lower *SI* value. The Root Mean Square (RMS) operator acts as a time collapsing function of the spatial information found in each of the video frames in order to obtain the parameter m_1.

The last two parameters m_2 and m_3 are measures of the video quality degradation due to losses of temporal information. The second parameter m_2 is defined as:

$$m_2 = f_{time}[0.108 * MAX(TI[O_n] - TI[D_n]), 0] \tag{5.2}$$

where

$$f_{time}(x_t) = STD_{time}CONV(x_t, [-1, 2, -1]) \tag{5.3}$$

The TI for the *nth* frame is the STD of the difference in pixel intensities between frame n and frame *n-1*. The operator TI is the temporal information lost between adjacent frames and is obtained by computing the STD of the difference in pixel intensity between adjacent frames. Video frames having little or no motion will have very low temporal information; while frames with high motion will have large temporal information values. *CONV* indicates the convolution operator is applied to the values corresponding to frames *n-1*, *n*, and *n+1* using the coefficients -1, 2 and -1, respectively. Since the convolution kernel is a high pass filter, the local motion measured by the m_2 parameter is enhanced. The function STD_{time} means the STD of the values computed by the convolution at each frame in the video sequence.

The third and final parameter, m_3 is another measure of temporal distortion. However, unlike the previous parameter, m_3 selects the frame which has the most added motion; it is defined as:

$$m_3 = MAX_{time}4.23 * LOG_{10}(TI[D_n]/TI[O_n]) \tag{5.4}$$

where MAX_{time} corresponds to the maximum value found among all of the frames. This parameter will give a heavier weighting to the frame with the most jerkiness or worst

uncorrected block; unlike m_2 where the worst block tended to be averaged out with the rest.

Least squares analysis was applied on 64 of the 132 clips (which comprised of 18 of the 36 test scenes used in training), to derive the coefficients in Equation 5.5 [37]; the remaining video sequences were used in testing.

$$\hat{s} = 4.77 - 0.992m_1 - 0.272m_2 - 0.356m_3 \tag{5.5}$$

where $\hat{s}$ is the estimated subjective quality assessment score of the video clip; we use this estimated score to evaluate the decoded video sequence quality.

A description of the implementation details pertaining to the video quality assessment algorithm follow. During the encoding of the video sequence in our experiments, while each frame is processed, the spatial and temporal information are computed for both the original and reconstructed frames. In order to determine the loss of spatial and temporal information due to video compression and possible frame losses, there must be a one-to-one mapping of the original video frames and those reconstructed after H.263 image compression is applied. Therefore, after a frame is encoded, the capacity of the buffer is checked to determine whether the frame will be discarded due to buffer overflow conditions, regardless of the video encoder algorithm used. If a frame is discarded, then the SI for the frame is set equal to that of the reconstruction of the last frame, because the previous frame will continue to be displayed at the receiver from video memory until the next frame is received. Similarly, when a frame is lost, the TI pertaining to that frame's reconstruction is set to 0, because there is no motion between the current and previous frame.

Once the entire video sequence has been computed, the arrays containing the spatial and temporal information for both the original and reconstructed video sequences are used to compute the parameters m_1, m_2, and m_3. These values are then used in the equation shown above to compute the video quality assessment rating.

Jitter

The video quality assessment algorithm just discussed, measures the degradation of the video quality caused by the loss of spatial and temporal information. However, when a low capacity network connection is used, large-sized frames will take a longer period of time to be transmitted causing the playback of the scenes at the decoder to exceed the 1/30th of a second inter-frame display time. If the inter-frame display time is excessive, the user may find this annoying since the video playback will appear jerky, this is known as jitter. Since

jitter is not detected by the video quality assessment algorithm we are using, the video quality in our experiments is determined by the rating of the video quality assessment algorithm and the amount of jitter present.

The magnitude of jitter present in a video sequence is computed as the standard deviation of the inter-frame display times [25]. The inter-frame delay is the amount of time that a frame spends in the buffer in excess of 1/30 of a second, since the frame rate used in our experiments is 30 frames/second. When a frame is transmitted within a single frame period, the inter-frame delay for the frame is 0 msec. Thus frames which are transmitted in less than 1/30th of a second do not add to the jitter; discarded frames are not included in our computation of jitter.

The MPEG-2 standard suggests jitter should be limited to ± 4 milliseconds [10]. Jitter outside this range can result in video degradations such as color deteriorations, audio breaks, or possibly frame freezing. We expect larger buffers will result in a greater ability to smooth the video traffic, increasing the utilization of the CBR connection. However, the inter-frame display will also become more variable causing the effects of jitter to become more apparent. Therefore, we will investigate the correlation between the buffer requirements, link utilization and the jitter.

5.2.2 Network Assessment Criteria

Network assessment criteria used in our experiments include the utilization of the CBR connection and the buffer requirements. Since the utilization and buffer requirements are affected by the burstiness of the bit stream, the mean and variance of the compressed frames sizes is also measured.

The 6 byte header overhead associated with the ATM and AAL 1 protocols decreases the real utilization of the CBR connection; that is the utilization of the connection based on the payload data only. The utilization discussed in our experiments includes the overhead of the cell headers. This utilization is used because a maximum utilization of 100% is more comprehensible than 88.7% (47 bytes/53 bytes) utilization computed using only the cell payload. A high utilization of the CBR connection is expected if buffer underflow conditions are avoided, since the buffer would always contain data available for transmission over the ATM network. However, buffer overflow should also be avoided as this will cause compressed video data to be discarded.

In our experiments we record the mean frame size and the corresponding variance as these parameters are expected to affect the transmission performance of compressed video

over a CBR ATM connection. This information will help users in the future to identify the CBR connection capacity required to transmit the VBR data and avoid exceeding the peak rate requested. We are expecting that the required CBR connection capacity will be the *mean bit rate * weighting_factor* where the weighting_factor (>1) will be affected by the variance in the VBR video stream; high motion video streams are expected to have a higher weighting_factor than video clips which contain low amounts of temporal information. A description of the video sequences used in our experiments is provided in Section 5.3.

Although larger buffers allow for better traffic shaping, they also cause the jitter to increase. We expect an inverse relationship between the CBR connection utilization and the corresponding jitter because low capacity connections would have a high utilization while the large-sized frame inter-frame display times would exceed 1/30th of a second. Alternatively, increasing the CBR connection capacity to reduce the jitter would reduce the utilization of the dedicated CBR connection.

5.3 Selection of Test Video Sequences

In this section we provide an analysis of the five video sequences which are available for testing. On the basis of our analysis we select two video sequences having the lowest and greatest amounts of spatial and temporal information, *i.e.*, the two sequences which have the most diverse amounts of information present.

Table 5.1 shows the amount of spatial and temporal information present in each of the five video sequences. Using the video quality assessment algorithm, the quantity of spatial information present in the original video sequence is computed using the standard deviation of each Sobel-filtered frame (as described in Section 5.2.1), an operation required for the computation of the m_1 parameter of the video quality assessment. The temporal information is computed using the standard deviation of the differences between the pixel luminance values computed from successive frames. The information in Table 5.1 is computed using the original images prior to any image compression. Therefore, the quantization parameter used and frequency of intra-encoding does not affect these values. A frame from each of the video sequences and a brief description of each sequence is provided below.

Figure 5.3 shows a frame from the Carphone video sequence. The large amount of spatial information present in the Carphone video sequence is attributed to the high amount of background detail which includes objects visible through the car window. The temporal information of this sequence, 6.7, is approximately half way between that of the Claire (2.2) and Foreman (10.0) sequences which respectively have the lowest and highest amounts of

Video Clip	Spatial Information Mean	Temporal Information Mean
Carphone	94.7	6.7
Claire	99.0	2.2
Foreman	102.0	10.0
Miss America	48.0	2.8
Salesman	74.3	3.3

Table 5.1: Spatial and temporal information present in the video sequences.

temporal information. The temporal information of the Carphone video sequence is caused by the large motion of the man's mouth and upper body, as well as the changing background viewed through the car window.

Figure 5.3: A frame in the Carphone video sequence.

Figure 5.4 depicts a frame from the Claire video sequence, which is a recording of a woman speaking some distance away from the camera. Although a large portion of each frame corresponds to the background, the spatial information is quite high (99.0) because of the varying background intensity. Since the woman is not near the camera and the motion is confined to her facial expressions and some movement of her head, the temporal information of this sequence is the lowest of the five video sequences analyzed.

Figure 5.5 illustrates a single frame from the Foreman video sequence. The Foreman video sequence has the highest amounts of spatial and temporal information of the five video sequences analyzed. The large amount of spatial information present is attributed to large amount of detail present in the foreground and background. The large amount of temporal information present is caused by large movements of the man's upper body as well as some movement of the camera causing the background to be shifted as well.

Figure 5.6 shows a frame from the Miss America video sequence. This sequence has the lowest amount of spatial information present and nearly the lowest amount of temporal

Figure 5.4: A frame in the Claire video sequence.

Figure 5.5: A frame in the Foreman video sequence.

information. Frames in this sequence contain a large amount of background with little information, most of the spatial information corresponds to the woman's facial features. Similarly, because most of the motion in the sequence is confined to the subject's face, the amount of temporal information present is small.

Figure 5.6: A frame in the Miss America video sequence.

Figure 5.7 shows a frame from the Salesman video sequence, the last sequence studied. The spatial information found in this sequence is approximately midway between the amounts found in the video sequences containing the smallest and largest amounts of spa-

tial information. Although there are a number of books and other objects present in the background, the area to the right of the speaker is fairly dark and contains little detail. The amount of motion in the video sequence is low and corresponds to movement of the man's mouth and arms while displaying the rectangular object held in his hands.

Figure 5.7: A frame in the Salesman video sequence.

As shown in Table 5.1, the Miss America video sequence has the least amount of spatial information present and almost the lowest temporal information; while the Foreman video clip has the most spatial and temporal information present. Separating the information encoded into its spatial and temporal components we see that the Foreman video has slightly more than twice (102/48) the spatial information when compared to the Miss America clip. The temporal information in the Foreman clip is nearly 4 times (10.0/2.8) that of the Miss America video clip.

When comparing the size of the P-frames (shown in Table 5.2) to the intra-encoded frames, we see the average P-frame size is about 19.9% (6239.6/31280.0) of the average I-frame size for the Foreman video sequence and about 9.9% (1384.2/14024.0) for the Miss America video sequence. Frame numbers 1 and 100 are intra-encoded. Since the mean P-frame size of the Foreman video clip is closer to that of their I-frame counterparts, the jitter is expected to be lower for this sequence than for the Miss America sequence when a low capacity CBR connection is used.

	148 P-frames		2 I-frames	
Video Clip	Mean	STD	Mean	STD
Foreman	6239.6	1354.1	31280.0	504.0
Miss America	1384.2	351.9	14024.0	384.0

Table 5.2: Frame sizes in the Foreman and Miss America video sequences.

Since the Miss America video sequence has the lowest amounts of spatial and temporal

information present, while the Foreman video sequence has the highest, these two sequences are used in our experiments. In our experiments, the video sequences are encoded using a quantization parameter of 7, to obtain comparable video quality values of slightly greater than 4.0. Having a video quality rating greater than 4.0 allows some flexibility for degradation, due to possible frame losses should buffer overflow occur, and still achieve a final quality assessment of at least 4.0. The corresponding bit rate for the Miss America sequence is about 47 kbps while the Foreman sequence yielded a much higher bit rate of about 197 kbps; an increase of approximately 300%.

Chapter 6

Experiments Using Intra-encoding at the Frame Level

This chapter discusses the experiments performed when entire frames are intra-encoded. In the experiments discussed in this chapter, frames 1 and 100 are intra-encoded, the remaining frames are inter-encoded as P-frames. The H.263 specification [11] recommends that each macroblock be intra-encoded at least once every 132 times in order to prevent error propagation [39]. Because each frame in QCIF consists of 99 macroblocks (9x11), we chose to encode every 99th frame instead of every 132nd in order to evaluate the performance when entire frames are intra-encoded versus when intra-encoding is done at the macroblock level (described in Chapter 7). The following section discusses the experiments run and the results obtained when no frames are discarded since buffer overflow is avoided. Section 6.2 provides an overview of the experiments run and the results obtained when frames are discarded due to buffer overflow.

6.1 Performance When No Frames are Discarded

This section gives an overview of the experiments run and the results obtained when no frames are discarded (that is, buffer overflow is not encountered). The video quality for the Foreman and Miss America video sequences when intra-encoding is performed at the frame level using a QP=7 and no frames are discarded is 4.318 and 4.295, respectively. Intra-encoding in the following experiments is done at the frame level and every 99th frame is intra-encoded. A description of the experiments follow:

- Experiment 1: Limited CBR Connection Capacity with Unlimited Buffering: the link utilization, buffer requirements and the corresponding jitter are measured when the CBR connection capacity is limited to the average bit rate of the video sequence.

Unlimited buffering is provided to prevent loss of video frames at the network source (Test 1, Section 6.1.1).

- Experiment 2: No buffering and Increased Capacity of the CBR Connection: the CBR connection capacity and utilization are measured when no buffering is required at the network source since the CBR connection capacity is increased to the point where the largest frame can be transmitted in a single frame period (Test 2, Section 6.1.2).

6.1.1 Limited CBR Connection Capacity, and Unlimited Buffering (Test 1)

In the first test, we simulate the transmission of the video sequences using a CBR connection capacity equal to the average encoded bit rate of the video sequence. Because the I-frame sizes are much larger than the overall average frame size, congestion occurs at the network source. Therefore, unlimited buffering is available in order to avoid any loss of frames. In this test we measure the link utilization, the maximum and mean buffer requirements, and the resulting jitter. These results are shown in Table 6.1 for the Foreman and Miss America video sequences.

	CBR Conn. Capacity	Link Util.	Buffer Requirements		Jitter
Video Sequence	(bits/sec)	(%)	Max.	Mean.	(msec)
Foreman	192128	99.0	45904	28145.2	10.7
Miss America	44149	97.1	29664	12425.4	24.1

Table 6.1: Test 1 results, limited CBR connection capacity and unlimited buffering.

In this table we see that the resulting jitter value is significantly lower for the Foreman clip (10.7 msec versus 24.1 msec); the jitter of the Foreman clip is approximately 44% that of the Miss America. The jitter is caused by the I-frames (frame numbers 1 and 100) which then delay all P-frames that are queued until the buffer is emptied out. However, since ratio of the mean P-frame size to the mean I-frame size, is larger for the Foreman clip than for Miss America (as discussed in Section 5.3), the I-frames do not delay as many P-frames at the network source, and for those frames that are delayed, the delays are shorter. This ratio also affects the buffer utilization. The mean buffer utilization for the Foreman and Miss America video clips is 61.3% and 41.9%, respectively. Although a buffer of unlimited size is used in this experiment, the mean buffer utilization is computed using the buffer requirements recorded in the experiment.

The mean utilization of the CBR connection, 99.0% for the Foreman video clip is greater

than the 97.1% utilization for the Miss America video sequence; this is caused by the variability of the P-frame sizes. Although buffer underflow does not occur when the Foreman video sequence is used, buffer underflow does occur with the Miss America sequence, since frame numbers 38-48 and 53-58 are sent before the next frame is added to the buffer. Because the Foreman frame sizes are more uniform than those of the Miss America video sequence, the Foreman produced substantially lower amounts of jitter and a higher link utilization.

6.1.2 No Buffering and Increased Capacity of the CBR Connection (Test 2)

This test simulates transmission of the video sequences over a CBR connection without any traffic smoothing at the network source. The buffer at the network is set just large enough to hold the largest I-frame and the CBR connection capacity is set so that the largest frame can be transmitted at the frame rate of 1/30th of a second. In this experiment we measure the CBR connection utilization, the results are shown in Table 6.2. Since the CBR connection capacity is large enough to transmit each frame during a single frame period the jitter is zero.

Comparing the results between the two video clips, we see that the Foreman clip obtained a link utilization of 21.3% which is significantly higher than the 10.8% link utilization obtained using the Miss America clip. Similarly, the CBR connection capacity for the Foreman clip had to be increased to 4.81 times that of its average bit rate; while the increase for the Miss America clip is 9.79 times its average bit rate to ensure that all the frames could be transmitted without any delays. The uniformity of the Foreman video sequence frame sizes results in a substantially higher utilization of the CBR connection when compared to the Miss America video sequence.

	Req'd CBR Connection Capacity	Link Utilization	Jitter
Video Sequence	bits/sec (× mean bit rate)	(%)	(msec)
Foreman	924136 (4.81)	21.3	0
Miss America	432219 (9.79)	10.8	0

Table 6.2: Test 2 results, no buffering and the CBR connection capacity is increased to the point where each frame is transmitted in a single frame period.

6.2 Performance When Frames are Discarded Due to Buffer Overflow

This section discusses the experiments run and the performance when frames 1 and 100 are intra-encoded and frames are discarded when buffer overflow is encountered. One use of the experiments in this section is to provide insight into whether it is better to discard small or large inter-encoded frames in a congested ATM network. We do not model congestion within the network because we are using a CBR quality of service. However, congestion within an ATM network may occur when a VBR quality of service is utilized or if the capacity of the CBR connection is exceeded. A description of each experiment is as follows:

- Experiment 1: The encoder is a black box: we monitor the CBR connection utilization, video quality, and jitter given various connection capacities (Test 3, Section 6.2.3).
- Experiment 2: A dynamic video encoder is used: we monitor the CBR connection utilization, video quality, and jitter given various connection capacities (Test 4, Section 6.2.4).
- Experiment 3: Discarding the Smallest and Largest Inter-encoded Frames: we monitor the video quality to identify which frame size should be assigned a higher priority in an ATM network (Test 5, Section 6.2.6).

The following section gives an overview of various frame discard algorithms. Section 6.2.2 discusses the computations used to determine the buffer sizes and CBR connection capacities used. Section 6.2.3 describes the experiments run when a black box video encoder is used and the results obtained, followed by a discussion of the results obtained when a dynamic video encoder is used in Section 6.2.4. Section 6.2.5 then compares the performance of the black box and dynamic video encoders. Finally, Section 6.2.6 discusses the results of an experiment run when only smallest and largest P-frames are discarded, to determine whether small or large-sized inter-encoded frames should be assigned a high priority in an ATM network.

6.2.1 Frame Discard Algorithms

This section gives an overview of the various potential methods of implementing frame discard algorithms and the benefits and drawbacks of each. Criteria to consider regarding which frames to discard when buffer overflow is encountered include: the *intelligence* of the encoder and the difficulty of the implementation. If the encoder is *intelligent*, we may

assume that when frames are assembled the application is aware of the buffer capacity. Then, if the buffer is full we can discard an entire frame, either the new frame to be added or one currently in the buffer. Otherwise, if the application has no knowledge of the buffer utilization, bits would be added on a first come first serve basis until the buffer is full, and the remaining bits of the frame would be discarded. A discussion of two algorithms for discarding a frame when buffer overflow is encountered is provided below.

The first possibility for discarding frames when the buffer is full is to discard the current frame to be added. Two alternatives to this approach include adding the portion of the frame that will fit in the available buffer space or discarding the entire frame. The latter assumes the upper layers are familiar with the buffer space available, which is applicable to our model, since buffering takes place above the ATM adaptation layer in the application layer. The former method will require more stringent buffer management since the frames will not be treated as integral blocks and will result in useless transmission of data across the network when bits pertaining to partial frames are sent.

A second possibility for discarding frames is to attempt to create buffer space by discarding frames thought to be less important (using heuristics) in order to provide space for the frame of higher priority. This frame prioritization would be in the form of discarding the frame(s) of lowest priority currently in the buffer. Low priority frames may be considered to be the oldest frames in order to reduce the jitter, or possibly the largest frames since this would alleviate congestion at the edge of the network the fastest. For example, if the buffer does not have the space required for the new frame to be added, the buffer would be searched and the largest frame discarded, thus freeing up as much buffer space as possible. However, this requires more processing. In addition, if part of the largest frame in the buffer has already been transmitted, the rest of the frame would be discarded, and the part of this frame transmitted would result in a waste of the CBR connection. Similarly, subsequent frames already in the buffer would be incorrectly decoded because the previous frame was lost. The implementation of this algorithm is complex and therefore is not used.

Since buffering is done at the application level, knowledge of the buffer occupancy is available to the video encoder. Therefore, due to the complexities outlined above, if the buffer capacity does not permit buffering of the entire current frame, then the current frame is discarded because partial frames are not stored in the buffer.

6.2.2 Computing the CBR Connection Capacities and the Buffer Sizes

This section provides an overview and motivation of the CBR ATM connection capacities

and the buffer sizes used in the following experiments. The mean bit rate is computed for each video sequence. The CBR connection capacity is then increased from the mean bit rate to the point where the largest I-frame can be transmitted in a single frame period. If the CBR connection capacity is near the mean bit rate, buffer overflow may occur causing a degradation of the video quality. The motivation for these experiments is to provide insight regarding the required capacity of the CBR connection in order to obtain acceptable jitter and video quality. Since the delay associated with the first frame is not included in the jitter computation, the size of the first I-frame is not included when computing the average bit rate. The equation used to derive the mean bit rate is as shown in Equation 6.1.

$$mean\ bitrate = ((total\ bits - first\ frame\ bits)/149\ frames) * 30\ frames/second \quad (6.1)$$

The CBR connection capacities (given as a multiple of the mean bit rate) that are used in the following experiments are as follows:

- Foreman CBR connection speeds: 1, 1.5, 2, 3, 4, 4.81
- Miss America CBR connection speeds: 1, 1.5, 2, 3, 4, 5, 6, 7, 8, 9, 9.79

The non-integral capacity of 1.5 is used due to the significant change in the magnitude of the results when the CBR connection capacity increased from the average bit rate to twice the average bit rate. The final connection capacities of 4.81 and 9.79 times the mean bit rate were chosen for the Foreman and Miss America video sequences, respectively, since these capacities facilitate the transmission of the largest I-frame (excluding the first frame) within a single frame period.

When the capacity of the CBR connection is low, *i.e.*, near the mean bit rate of the 149 frames, congestion in the buffer is expected particularly when an I-frame is added to the buffer. Therefore we implemented a *warning state* in the buffer which enables some flexibility regarding which frames should be discarded based on their size. Intra-encoded frames are important to prevent further propagation of errors found in P-frames and therefore are never discarded. Using the buffer warning state the largest or the smallest P-frames are discarded as a means of alleviating congestion at the buffer. These results are then used to determine whether the small- or large-sized P-frames are most important to keep in order to achieve a high video quality. This information is then used to determine which P-frame sizes should be assigned a high priority in an ATM network.

The buffer size in the following experiments is set to $I + 2\bar{P}$ where I is the size of the largest I-frame and $\bar{P}$ is the mean of the P-frames plus 2 standard deviations. This buffer

size ensures the largest I-frame can always be admitted to the buffer; while the additional 2 P-frames provide an opportunity to discard inter-encoded frames based on their size. A buffer *warning state* is encountered if there are more than 0 bits and less than the number of bits corresponding to 2 mean sized P-frames in the buffer. If the buffer contains more than the number of bits corresponding to 2 mean sized P-frames when a P-frame is about to be added to the buffer, the frame is discarded. This ensures I-frames are not discarded since they are crucial to preventing any further propagation of errors.

A threshold of 1 STD from the mean P-frame size is used to determine whether a frame should be discarded due to its size when buffer overflow is encountered. When a frame is to be admitted to the buffer, if the warning state is set, the frame is discarded in accordance with the discard criteria specified. For example, if the buffer warning state is set and the small inter-encoded frames are being discarded and this frame is more than 1 STD less than the mean P-frame size, the frame is discarded. Similarly, if the warning state is set and large-sized frames are being discarded and this frame is more than 1 STD larger than the mean, the frame is discarded. If a frame is to be added to the buffer but the number of bits in the buffer exceed the number of bits corresponding to 2 mean sized P-frames (*i.e.* an I-frame is in the buffer), the frame is discarded in order to prevent circumstances occurring which would prevent entire I-frames from being added to the buffer.

6.2.3 Results When a Black Box Video Encoder is Used and Buffer Overflow is Experienced (Test 3)

In this section we evaluate the performance when the *black box* encoder (discussed in Chapter 4) is used to encode the video sequence. Using the black box video encoder, additional errors may be introduced causing the video quality to be reduced when frames are discarded due to buffer overflow, since the inter-encoded frames reference a different frame than they were encoded with. This experiment is also used to judge the performance when intra-encoding is performed at the frame level based on the resulting jitter and CBR connection utilization.

The previous two tests (discussed in Sections 6.1.1 and 6.1.2) simulated transmission of the compressed video when the CBR connection capacity was equal to the average bit rate and when the CBR connection capacity was increased to the point where buffering at the network source was not required. Neither of these tests led to degradation of the video quality using our assessment algorithm since no frames were lost, although the former test did introduce a significant amount of jitter.

In Table 6.3, the link utilization, jitter, and the video quality are given for the different CBR connection capacities when the small- and large-sized P-frames are discarded for the Foreman video sequence. The number of frames discarded resulting from buffer overflow is also shown. The buffer size used in this experiment is 49.9 kbits. The capacity of the CBR connection is varied from the mean encoded bit rate of 192 kbps to 924 kbps which allows the largest frame (excluding the first I-frame) to be transmitted in 1/30th of a second. Excluding the first frame, frame number 100 is the largest frame and contains 30,776 bits. The numbers in parentheses in the link capacity column indicate the ratio of the link capacity to the mean bit rate.

	Discard Small-sized Frames				Discard Large-sized Frames			
Link Capacity (bits/sec)	Link Util. (%)	Jitter (msec)	Frames Disc'd	Video Quality	Link Util. (%)	Jitter (msec)	Frames Disc'd	Video Quality
192128 (1.0)	92.8	11.3	15	3.758	86.7	11.3	19	3.560
288192 (1.5)	67.3	6.1	2	4.182	66.2	6.1	4	3.848
384256 (2.0)	50.5	3.8	2	4.182	50.1	3.9	3	4.017
576384 (3.0)	34.0	1.6	1	4.295	33.8	1.6	1	4.189
768512 (4.0)	25.5	0.5	1	4.295	25.6	0.5	0	4.318
924136 (4.81)	21.3	0.0	0	4.318	21.3	0.0	0	4.318

Table 6.3: Test 3 results, Foreman video sequence when various CBR connection capacities are used with a black box encoder.

Analysis of the results in Table 6.3 indicate that as the CBR connection capacity is increased from the mean bit rate, the resulting jitter and number of bits lost decrease while the video quality increases. In order to obtain a video quality of at least 4.0 and a tolerable amount of jitter ($\leq$ 4 msec) for the Foreman video sequence, a CBR connection capacity of at least 2.0 times the mean bit rate is required using either the small- or large-sized frame discard criterion. Because the intra-encoded frames 1 and 100, are very large compared to the inter-encoded frames (see Table 5.2), the utilization of the CBR connection is less than 51% in the range where the jitter and video quality values are acceptable.

In Figure 6.1 the video quality is plotted against the connection capacity for the Foreman video sequence encoded with a black box video encoder. The graph shows the video quality when the small and large frames are discarded due to buffer overflow. As the graphs illustrate, it is unclear whether small- or large-sized inter-encoded frames should be assigned a high priority in an ATM network. Because the warning state lasts for a short duration, frames are often discarded not because the frame is very small or very large, but because the

buffer occupancy is beyond the warning state (the number of bits in the buffer exceeds the number corresponding to 2 mean-sized P-frames) which usually happens when an I-frame is added to the buffer and the CBR connection capacity is low. As a result, the lines in the graph cross. In order to fairly evaluate whether small or large frames are most important, a comparison should be performed when a similar number of bits are transmitted. Section 6.2.6 analyzes the results obtained from an experiment when only the smallest and largest P-frames are discarded and a similar number of bits are transmitted.

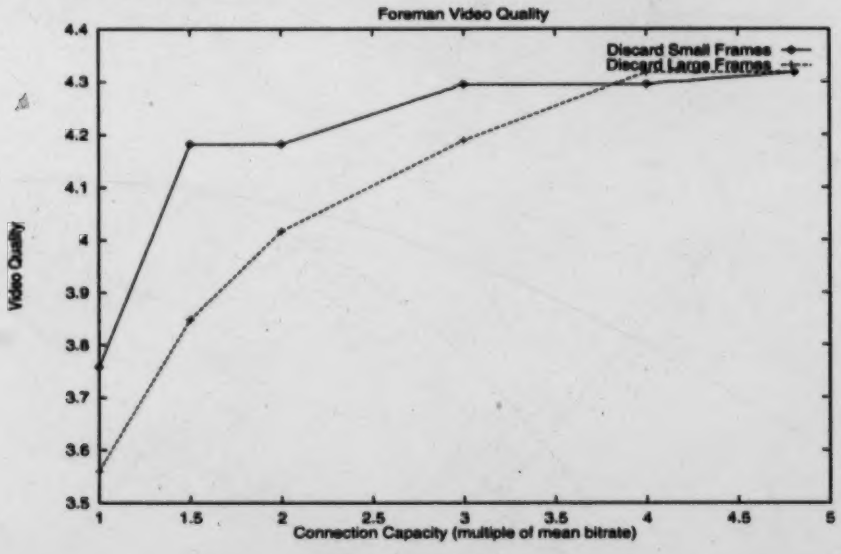

Figure 6.1: Video quality of the Foreman video sequence when a black box encoder is used and the small- and large-sized frames are discarded.

Table 6.4 shows the results when the CBR connection capacity is varied from the mean bit rate (44 kbps) to 9.79 times the mean bit rate (432 kbps) when the small-sized and large-sized P-frames are discarded for the Miss America video sequence. Increasing the capacity to 9.79 times the mean bit rate is required so that the largest frame (14,408 bits) can be transmitted in a single frame period. The buffer size used in this experiment is 18.5 kbits. Analysis of the results in this table are similar to those obtained when the Foreman video sequence was tested. Again, increasing the CBR connection capacity decreases the link utilization, jitter, and number of frames discarded due to buffer overflow, while the video quality continues to increase. In order for the Miss America video sequence to achieve a video quality of at least 4.0, and a jitter not exceeding 4 msec, a CBR connection capacity of 4 times the mean bit rate is required. This relatively large CBR connection capacity

results in a low network utilization of 26%.

	Discard Small-sized Frames				Discard Large-sized Frames			
Link Capacity (bits/sec)	Link Util. (%)	Jitter (msec)	Frames Disc'd	Video Quality	Link Util. (%)	Jitter (msec)	Frames Disc'd	Video Quality
44149 (1.0)	88.1	27.1	26	3.480	82.4	26.8	30	3.360
66224 (1.5)	66.1	15.6	9	3.607	65.6	15.7	10	3.602
88298 (2.0)	51.0	10.9	5	3.757	51.0	10.9	5	3.757
132447 (3.0)	34.5	6.3	3	3.874	34.5	6.3	3	3.874
176596 (4.0)	26.0	4.0	2	4.074	26.0	4.0	2	4.074
220745 (5.0)	21.0	2.7	1	4.157	21.0	2.7	1	4.157
264894 (6.0)	17.4	1.7	1	4.157	17.4	1.7	1	4.157
309043 (7.0)	15.0	1.1	1	4.157	15.0	1.1	1	4.157
353192 (8.0)	13.2	0.6	0	4.295	13.2	0.6	0	4.295
397341 (9.0)	11.7	0.2	0	4.295	11.7	0.2	0	4.295
432219 (9.79)	10.8	0.0	0	4.295	10.8	0.0	0	4.295

Table 6.4: Test 3 results, Miss America video sequence when various CBR connection capacities are used with a black box video encoder.

In Figure 6.2 the video quality is plotted against the connection capacity for the Miss America video sequence encoded with a black box video encoder, when the small- and large-sized frames are discarded. Again, frames are often discarded because the buffer is beyond the warning state, causing a P-frame to be discarded regardless of its size. When the connection capacity is two times the mean bit rate, or greater, the video quality using either discard criterion is the same since the same frames are discarded.

6.2.4 Results When a Dynamic Video Encoder is Used and Buffer Overflow is Experienced (Test 4)

In this section experiments are run to evaluate the performance when a dynamic video encoder is used to encode the Foreman and Miss America video sequences. The CBR connection capacities and buffer sizes used in this experiment are the same as those discussed in the previous experiment. The dynamic video encoder we use has the ability to monitor the buffer occupancy, and is aware when frames are discarded due to buffer overflow at the network source. This functionality permits the encoder to inter-encode the current frame using the reconstruction of the last frame transmitted. This in turn reduces the adverse effect on the video quality when a frame is discarded.

As discussed in the previous experiment, because the warning state lasts for a short duration it is unclear whether the small- or large-sized frames should be discarded. There-

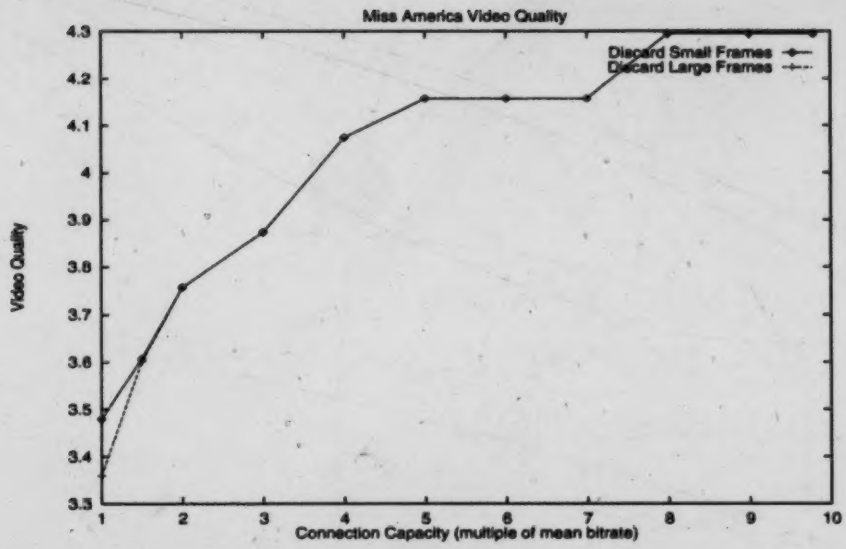

Figure 6.2: Video quality of the Miss America video sequence when a black box encoder is used and the small- and large-sized frames are discarded.

fore, in this section, the results obtained when the large frames are discarded have been omitted. A performance comparison based on the video quality produced by the black box encoder (discussed in the previous section) and the dynamic video encoder is provided in the following section.

Table 6.5 shows the connection capacities, utilization, jitter, the number of frames discarded, and the video quality when the Foreman video sequence is encoded using a dynamic video encoder. Analysis of the results show that as the CBR connection capacity is increased, the resulting jitter and number of bits lost decrease while the video quality increases. Again, in order to obtain a video quality of at least 4.0 for the Foreman video sequence, and a low amount of jitter ($\leq$ 4 msec), a CBR connection capacity of at least two times the mean bit rate is required when either the small or large frames are discarded. Due to the intra-encoded frames 1 and 100, which are very large compared to the inter-encoded frames (see Table 5.2), the utilization of the CBR connection is less than 51% in the range where the jitter and video quality values are acceptable for the Foreman video sequence.

Table 6.6 shows the results obtained when the dynamic video encoder is used to encode the Miss America video sequence. As discussed with the Foreman video sequence, increasing the CBR connection capacity reduces the jitter and number of bits lost, while the video quality increases. A link capacity of four times the mean bit rate is required to reduce the

Link Capacity (bits/sec)	Link Util. (%)	Jitter (msec)	Frames Discarded	Video Quality
192128 (1.0)	93.8	11.6	17	4.100
288192 (1.5)	67.3	6.1	2	4.209
384256 (2.0)	50.5	3.8	2	4.209
576384 (3.0)	34.0	1.6	1	4.289
768512 (4.0)	25.5	0.5	1	4.289
924136 (4.81)	21.3	0.0	0	4.318

Table 6.5: Test 4 results, Foreman video sequence when various CBR connection capacities are used with a dynamic video encoder and small-sized frames are discarded.

jitter to an acceptable level, the same rate required when the black box encoder is used.

Link Capacity (bits/sec)	Link Util. (%)	Jitter (msec)	Frames Discarded	Video Quality
44149 (1.0)	88.9	27.3	30	4.271
66224 (1.5)	66.8	15.7	10	4.313
88298 (2.0)	51.3	10.9	5	4.296
132447 (3.0)	34.6	6.3	3	4.324
176596 (4.0)	26.2	4.0	2	4.316
220745 (5.0)	21.0	2.7	1	4.301
264894 (6.0)	17.5	1.7	1	4.301
309043 (7.0)	15.0	1.1	1	4.301
353192 (8.0)	13.2	0.6	0	4.295
397341 (9.0)	11.7	0.2	0	4.295
432219 (9.79)	10.8	0.0	0	4.295

Table 6.6: Test 4 results, Miss America video sequence when various CBR connection capacities are used with a dynamic video encoder and small-sized frames are discarded.

The results of this test indicate that the required capacity of the CBR connection must be increased substantially in order to reduce the jitter to an acceptable value. The Foreman video sequence requires the connection capacity to be twice the mean bit rate when using either the black box or dynamic video encoder. The capacity of the CBR connection must be increased to four times the mean bit rate for the Miss America video sequence when either video encoder is used.

6.2.5 Performance Comparison of Black Box and Dynamic Video Encoding

In this section we compare the performance of our black box and dynamic video encoders based on the results obtained in Tests 3 and 4, respectively (Sections 6.2.3 and 6.2.4). The

jitter and utilization of the CBR connection are similar when either video encoder is used as described in the previous section. The results discussed in this section correspond to the experiments run when the small-sized frames are discarded; these results are similar to those obtained when the large-sized frames are discarded.

Table 6.7 shows the results when the black box and dynamic encoders are used to encode the Foreman video sequence. As shown in the table, using the dynamic video encoder to encode the Foreman video sequence produces a video quality greater than 4.0 even when the available CBR connection capacity is equal to the mean bit rate. However, using the black box encoder, a video quality rating exceeding 4.0 is not achieved until the CBR connection capacity is increased to 1.5 times the mean bit rate. When the connection capacity is near the mean bit rate, the dynamic video encoder produces a higher quality video than that of the black box encoder. This is due to the the dynamic video encoder's ability to avoid the errors that occur when a black box encoder is used. These errors are a result of the black box encoder's inability to monitor when a frame is discarded due to buffer overflow; this in turn causes frames to be decoded using a different frame than the one referenced during encoding. The video quality is further degraded because the errors are allowed to propagate since successive inter-encoded frames also reference these incorrectly decoded frames.

	Black Box Video Encoder			Dynamic Video Encoder		
Link Cap. (bits/sec)	Frames Discarded	Bits Discarded	Video Quality	Frames Discarded	Bits Discarded	Video Quality
192128 (1.0)	15	84728	3.758	17	104120	4.100
288192 (1.5)	2	11184	4.182	2	11184	4.209
384256 (2.0)	2	11184	4.182	2	11184	4.209
576384 (3.0)	1	3536	4.295	1	3536	4.289
768512 (4.0)	1	3536	4.295	1	3536	4.289
924136 (4.81)	0	0	4.318	0	0	4.318

Table 6.7: Performance comparison of the black box and dynamic video encoders using the Foreman video sequence.

When the CBR connection capacity is equal to the mean bit rate, the black box encoder discards fewer frames than the dynamic encoder because the loss of a frame when dynamic encoding is used increases the size of the next frame encoded. The larger frame size is due to more information being encoded since the frame referenced for inter-encoding is the reconstruction of the last frame added to the buffer rather than the last frame encoded as described in Chapter 4. Although the dynamic encoder may discard more frames due to buffer overflow, it still provides a higher video quality when many frames are discarded.

In Figure 6.3, the video quality is plotted against the connection capacity, when both the black box and dynamic video encoders are used to encode the Foreman video sequence. As shown in the figure, the performance of the dynamic video encoder provides a substantial improvement in the resulting video quality compared to when the black box encoder is used and the available CBR connection capacity is low. When the CBR connection is increased to the point where no frames are discarded, both the black box and dynamic video encoders produce the same video quality.

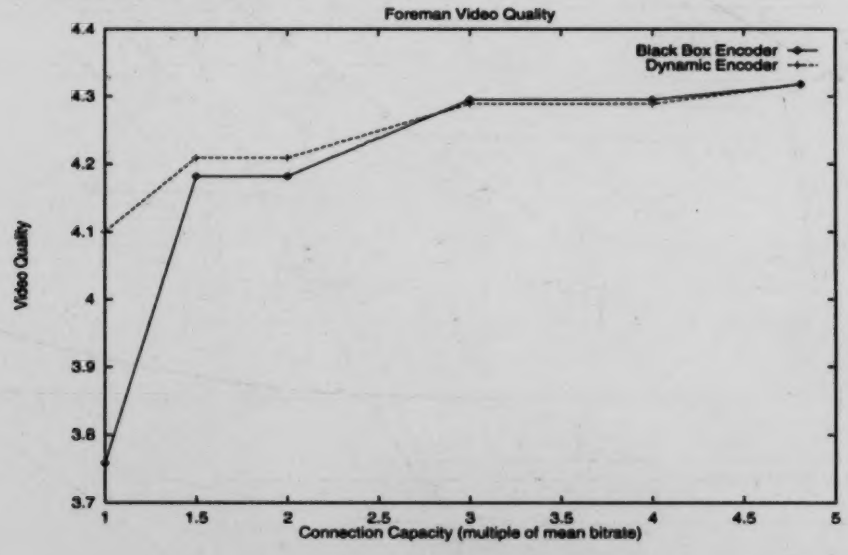

Figure 6.3: Video quality of the Foreman video sequence using both a black box and dynamic video encoder when the small- or large-sized frames are discarded.

For completeness, the results obtained when a dynamic video encoder is used to encode the Miss America video sequence are included in Table 6.8. These results are similar to the results discussed when the Foreman video sequence was tested. That is, the dynamic video encoder provides a video quality rating of more than 4.0 even when the CBR connection capacity used is equal to the mean bit rate. However, the black box encoder does not produce a video quality of more than 4.0 until the connection capacity is four times the mean bit rate. Again, discarding a frame when the dynamic encoder is used causes the next frame encoded to become larger. This in turn causes a slight increase in the number of frames discarded compared to when the black box encoder is used and the CBR connection capacity is low, *i.e.*, 30 instead of 26 frames are discarded when the connection capacity is equal to the mean bit rate.

	Black Box Video Encoder			Dynamic Video Encoder		
Link Cap. (bits/sec)	Frames Discarded	Bits Discarded	Video Quality	Frames Discarded	Bits Discarded	Video Quality
44149 (1.0)	26	38560	3.480	30	58080	4.271
66224 (1.5)	9	13592	3.607	10	19464	4.313
88298 (2.0)	5	7072	3.757	5	8520	4.296
132447 (3.0)	3	4296	3.874	3	4664	4.324
176596 (4.0)	2	2608	4.074	2	2608	4.316
220745 (5.0)	1	1432	4.157	1	1432	4.301
264894 (6.0)	1	1432	4.157	1	1432	4.301
309043 (7.0)	1	1432	4.157	1	1432	4.301
353192 (8.0)	0	0	4.295	0	0	4.295
397341 (9.0)	0	0	4.295	0	0	4.295
432219 (9.79)	0	0	4.295	0	0	4.295

Table 6.8: Performance comparison of the black box and dynamic video encoders using the Miss America video sequence.

In Figure 6.4, the video quality is plotted against the connection capacity for the Miss America video sequence encoded with the black box and dynamic video encoders. Again, this figure illustrates that the dynamic video encoder provides superior video quality than obtained using a black box encoder, particularly when the CBR connection capacity is near the mean bit rate.

6.2.6 Discarding the Smallest and Largest Inter-encoded Frames (Test 5)

This experiment, Test 5, is used to obtain results when only the smallest or largest P-frames are discarded, and a similar number of bits are transmitted. Since the frame discard criteria used in Tests 3 and 4 usually resulted in a different number of frames being discarded, and more importantly a different number of bits being discarded, a fair comparison could not be made. The results from this experiment will allow us to determine whether the small- or large-sized inter-encoded frames should be assigned a higher priority in an ATM network.

We use the dynamic video encoder in the following experiments since the black box encoder often produces a poorer video quality, particularly when a large number of frames are discarded, as discussed in the previous section. The threshold used to determine whether a frame's size fits the discard criterion is found by discarding all the small-sized frames whose size is more than 1 STD less than the mean. Using the number of bits transmitted when the small frames are discarded, we adjust the number of standard deviations until the largest

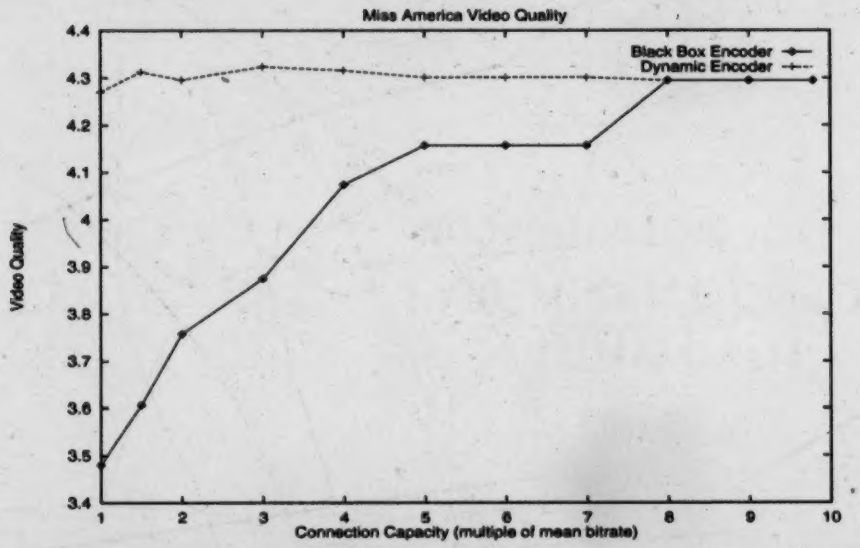

Figure 6.4: Video quality of the Miss America video sequence using both a black box and dynamic video encoder when the small- or large-sized frames are discarded.

frames are discarded and approximately the same number of bits are transmitted. The large-sized frames in the Foreman video sequence correspond to those frames whose size is more than 1.68 standard deviations greater than the mean frame size, and 1.49 standard deviations greater than the mean for the Miss America video sequence.

Table 6.9 shows the results when the smallest and largest frames of the Foreman and Miss America video sequence are discarded. The numbers in parenthesis in the frame "discard criteria" column indicate the number of standard deviations from the mean used to determine whether a frame's size is considered *small* or *large*. As the results show, the difference in video quality degradation when either small- or large-sized frames are discarded is small.

Video Sequence	Discard Criteria	Frames Discarded	Bits Transmitted	Video Quality
Foreman	Small Frames (1.0)	15	942512	4.163
Foreman	Large Frames (1.68)	7	940768	4.116
Miss America	Small Frames (1.0)	23	217312	4.239
Miss America	Large Frames (1.49)	10	216432	4.243

Table 6.9: Test 5 results, when only the smallest and largest P-frames are discarded.

Table 6.10 shows the video quality for each sequence when no frames, small frames and

then large frames are discarded. The data in this table is included to provide a breakdown of the video quality shown in Table 6.9. The video quality, when no frames are discarded is included in Table 6.10 to illustrate how much the video quality is reduced when either the small or large frames are discarded. In addition, the three parameters (m_1, m_2, and m_3) used to compute the resulting video quality are also included. The first parameter, m_1, which measures the loss of spatial information is largest when the small frames are discarded because more small-sized frames are discarded than large-sized frames. The second parameter, m_2, measures the overall loss of temporal information and is greatest when the large-sized frames are discarded. Discarding the large-sized frames results in a larger loss of temporal information because the large-sized frames correspond to periods of higher motion. The third parameter, m_3, selects the frame with the most added motion, and is greatest when the small-sized frames are discarded. Since this parameter is computed using a ratio of the temporal information found in the decoded and original video sequences, it is more sensitive to the loss of small-sized frames. The Foreman video sequence which has larger amounts of motion (as shown in Table 5.1) suffers a greater degradation of the video quality, nearly 5% (4.116/4.318) in the worst case compared to the video quality degradation experienced in the Miss America video sequence, approximately 1.5% (4.239/4.295).

Video Sequence	Discard Criteria	m_1	m_2	m_3	Video Quality
Foreman	None	0.237	0.047	0.574	4.318
Foreman	Small Frames	0.239	0.444	0.702	4.163
Foreman	Large Frames	0.233	0.775	0.597	4.116
Miss America	None	0.322	0.039	0.407	4.295
Miss America	Small Frames	0.331	0.168	0.442	4.239
Miss America	Large Frames	0.317	0.249	0.407	4.243

Table 6.10: Video quality analysis when the smallest and largest P-frames are discarded.

The results from this experiment indicate that video sequences comprised of low motion (*e.g.*, the Miss America sequence) have a similar video quality degradation regardless of whether small or large frames are discarded. However, video sequences containing larger amounts of motion, such as the Foreman video sequence, should avoid discarding large-sized frames since this results in more serious losses of temporal information. Therefore, although the loss of frames should be avoided as much as possible, large-sized frames should be assigned a high priority in an ATM network based on the results obtained using the Foreman and Miss America video sequences.

The experiments up to this point have intra-encoded frames 1 and 100, and the remaining

148 frames were encoded as P-frames. Since the I-frames are much larger than the P-frames they decrease the utilization of the CBR connection and increase the corresponding jitter. The short traces we are using consist of 150 frames and constitute a limited sample. Rather than rigorously investigating the results using larger traces, we consider a new direction (inter-encoding at the macroblock level) which promises to solve some of the problems associated with transmitting the large I-frames over a CBR ATM connection. This algorithm is described in the following chapter.

Chapter 7

Experiments Using Intra-encoding at the Macroblock Level

In the previous chapter we showed that intra-encoding at the macroblock level results in poor network performance. In order to reduce the jitter to an acceptable level, the utilization of the CBR connection for the Foreman and Miss America video sequences is approximately 51% and 26% respectively. Since frames 1 and 100 are intra-encoded they significantly reduce the network utilization, and increase the buffer requirements at the network source resulting in larger jitter. Therefore, in order to achieve higher network link utilization and lower jitter values when the available CBR connection capacity is near the mean bit rate, we propose intra-encoding at the macroblock level. This is in accordance with the H.263 recommendation [11] which indicates each macroblock should be intra-encoded at least once every 132 times. Intra-encoding at the macroblock level has been investigated before as a means of increasing the robustness of H.263 encoded video when the network connection is Internet [39] (discussed in Chapter 1). However, in our work we investigate this encoding algorithm as a means of increasing the utilization of our CBR ATM connection.

This chapter provides an overview of our macroblock intra-encoding algorithm and the results obtained when intra-encoding is done at the macroblock level. The experimental design is the same as described in Section 5.1. The experiments are:

- Experiment 1: Performance When the Frequency of Macroblocks Intra-encoded is Varied: monitor the bits generated and the resulting video quality when a various number of macroblocks are intra-encoded in P-frames (Test 6, Section 7.2).
- Experiment 2: Various CBR Connection Capacities with Unlimited Buffering: monitor the CBR connection capacity, link utilization, jitter, and buffer requirements (Test 7, Section 7.3).

The following section provides an overview of our macroblock intra-encoding algorithm. In Section 7.2, we analyze the results obtained when the Foreman video sequence is encoded and the number of macroblocks intra-encoded in each frame is varied. A discussion of the results obtained using the Foreman and Miss America video sequences when a single macroblock is intra-encoded in each of the 149 P-frames following the first I-frame is provided in Section 7.3. In Section 7.4, a performance comparison is done based on the results obtained when the Foreman and Miss America video sequences are intra-encoded at the frame and macroblock levels. The performance of intra-encoding at the macroblock level is further evaluated, in Section 7.5, using the Carphone, Claire and Salesman video sequences. Finally, in Section 7.6, a summary is provided based on the results obtained when each of the five video sequences is intra-encoded at the macroblock level.

7.1 Macroblock Intra-encoding Algorithm

This section provides an overview of the macroblock intra-encoding algorithm used to alleviate the problems associated with intra-encoding entire frames. These problems include: poor utilization of the CBR connection, large amounts of jitter, and reductions of the video quality when frames are discarded due to buffer overflow. The cause of these problems is that the intra-encoded frames are several times larger than their inter-encoded counterparts. For the Foreman video sequence, the average I-frame is 5 times larger than the average-sized P-frame, this ratio increases to 10 for the Miss America video sequence (as shown in Table 5.2). We intend to eliminate the problems caused by these large I-frames by introducing intra-encoding at the macroblock level, now only the first frame is intra-encoded followed by 149 P-frames.

Intra-encoding at the macroblock level required some modifications to the encoder in order to specify which macroblocks of a P-frame would be intra-encoded; the original implementation simply inter-encoded all the macroblocks of a P-frame. Our selection of macroblocks to be intra-encoded is cyclic; that is, for example, if each frame is to have two macroblocks intra-encoded, the first frame would have the first and second macroblocks intra-encoded, the second frame would have the third and fourth macroblocks intra-encoded and so on. This method was chosen for its ease of implementation and in order to evaluate the performance of this method. More complex methods for selecting the macroblock to be intra-encoded in order to further enhance the video quality are suggested in Section 8.2, which outlines future research.

Our experiments utilize the Quarter Common Intermediate Format (QCIF) which has a

resolution of 144 lines by 176 pixels (or 9 by 11 macroblocks) and contains 99 macroblocks per frame. Therefore, in the following experiment we vary the number of macroblocks intra-encoded per frame as follows: 1, 3, 9, 11, 33, and 99. These numbers were chosen as they are integer divisors of 99, and they ensure that each macroblock is intra-encoded at least once every 132 frames in compliance with the H.263 recommendation. Although intra-encoding more frequently is expected to increase the resulting bit rate, it should decrease the propagation of errors caused by the precision mismatches of the DCT operations and increase the resulting video quality [39].

7.2 Performance When the Frequency of Macroblocks Intra-encoded is Varied (Test 6)

In this section we evaluate the performance when the frequency of macroblocks intra-encoded is varied. Therefore, we adjust the number of macroblocks that are intra-encoded and monitor the resulting video quality and the number of bits generated in order to evaluate the benefits of more frequent intra-encoding. No frames are lost in this experiment so the resulting video quality is strictly due to the number of macroblocks intra-encoded. Using the Miss America and Foreman video sequences the number of macroblocks (per frame) that are intra-encoded is varied using integer divisors of 99, from 0 to 99. When all 99 macroblocks of each P-frame are intra-encoded, this causes the first frame (which is always intra-encoded) to be followed by 149 I-frames.

Table 7.1 shows the results when the number of macroblocks intra-encoded in the Foreman video sequence is varied. The first column indicates the number of macroblocks intra-encoded in each of the 149 P-frames, and the second column shows the total number of bits generated for the 150 frame sequence. In addition, the mean, standard deviation and coefficient of variation for the P-frame is also shown, as well as the three parameters used to compute the video quality. In each of these runs, the size of the first frame for the Foreman video sequence remains constant, 31,784 bits. Definite trends are evident in this table that correspond to the increasing frequency of the number of macroblocks intra-encoded. The most obvious is that as the number of intra-encoded macroblocks increases, so does the number of bits required to encode the frame. In addition, as the number of macroblocks intra-encoded in each frame increases, the coefficient of variation decreases, indicating less variability among the inter-encoded frame sizes. This is due to the reduction in the amount of temporal redundancy being encoded.

Although the video quality does increase when the number of macroblocks intra-encoded

MBs (intra)	Total Bits	P-frame			VQ	m_1	m_2	m_3
		Mean	STD	Coef. Var.				
0	960280	6232	1349	0.05	4.327	0.244	0.052	0.524
1	997456	6481	1364	0.04	4.356	0.234	0.050	0.475
3	1076016	7008	1409	0.04	4.338	0.222	0.045	0.560
9	1300768	8517	1449	0.03	4.312	0.192	0.042	0.720
11	1378352	9037	1399	0.02	4.301	0.192	0.046	0.749
33	2239848	14819	1771	0.01	4.363	0.139	0.032	0.730
99	5064048	33774	1280	0.00	4.417	0.074	0.009	0.780

Table 7.1: Intra-encoding the Foreman video sequence at the macroblock level.

is increased, we see a continuous decrease in the amount of spatial detail lost (which is measured by the parameter m_1). Similarly, the overall loss of temporal information measured by the parameter m_2 continues to decrease. The third parameter m_3 which captures the frame with the largest added motion generally increases with the frequency of intra-encoded macroblocks. It is important to note that when all of the macroblocks (99 for the QCIF encoding format used here) are intra-encoded the highest video quality occurs. However, comparing the cases when only a single macroblock versus when all 99 macroblocks are intra-encoded, the video quality increases by only 1.5% while the number of bits generated increased by over 400%.

Table 7.2 shows the results for the Miss America video sequence when the number of macroblocks intra-encoded is varied. The number of macroblocks intra-encoded and the number of bits generated are shown in the first and second columns, respectively. The table also shows the statistical properties of the P-frames and the video quality that results when the frequency of macroblocks intra-encoded changes. The size of the first frame, the only I-frame in each of these runs, is 13,640 bits. The trends observed in Table 7.2 for the Miss America video sequence are similar to those discussed for the Foreman video sequence. However, since the Miss America video sequence contains much less motion than the Foreman video sequence (see Table 5.1) the coefficient of variation changes very little for the Miss America clip until finally every macroblock of each frame is intra-encoded. Again when all of the macroblocks are intra-encoded compared to when a single macroblock per frame is intra-encoded, a marginal increase (4.5%) in video quality occurs, while the number of bits generated increases substantially; by more than 800%!

As shown in Tables 7.1 and 7.2, increasing the frequency of intra-encoded macroblocks marginally increases the resulting video quality at a substantial cost in terms of the much

MBs (intra)	Total Bits	P-frame			VQ	m_1	m_2	m_3
		Mean	STD	Coef. Var.				
0	222272	1400	370	0.07	4.286	0.331	0.038	0.408
1	238928	1512	386	0.07	4.300	0.320	0.034	0.402
3	278616	1778	443	0.06	4.321	0.294	0.031	0.419
9	387528	2509	561	0.05	4.291	0.241	0.029	0.652
11	425544	2765	576	0.04	4.329	0.231	0.030	0.573
33	843112	5567	1395	0.06	4.253	0.146	0.030	1.023
99	2170648	14477	203	0.00	4.492	0.046	0.004	0.649

Table 7.2: Intra-encoding the Miss America video sequence at the macroblock level.

larger resulting bit rate. Since our intent is to transmit the compressed video stream over a CBR connection using minimal network resources while maintaining a satisfactory video quality, we choose to intra-encode 1 macroblock per frame.

7.3 Various CBR Connection Capacities with Unlimited Buffering (Test 7)

The previous experiments have shown that intra-encoding at the frame level (*i.e.*, every 99th frame) results in low network utilization, poor video quality and undesirably high jitter values. Therefore, since intra-encoding at the macroblock level creates much less variance in the resulting compressed video frame sizes and is in accordance with the H.263 video standard, this algorithm is expected to perform better. In this test, the CBR connection capacity is varied from the mean bit rate of the inter-encoded frames up to the value where no buffering is required because the largest frame can be transmitted before the next frame is encoded, corresponding to the inter-frame display time of 1/30 second.

Previous tests have shown that losses of inter-encoded frames may significantly degrade the resulting video quality, therefore no losses are allowed in this experiment. Since intra-encoding is performed at the macroblock level instead of at the frame level, the buffer requirements are not as great because all frames excluding the first frame (which is an I-frame) are much more uniform in size.

The following section discusses the computations used to determine the capacity of the CBR connection when intra-encoding is performed at the macroblock level. Section 7.3.2 shows the results obtained using the Foreman and Miss America video sequences when intra-encoding is performed at the frame level.

7.3.1 Computing the CBR Connection Capacities

This section shows the computations used to calculate the minimum capacity of the CBR connection used in these experiments. Once the minimum capacity (corresponding to the mean bit rate of the P-frames) has been determined, the capacity of the CBR connection is then increased in increments of 5% until the jitter is reduced to 0.

Encoding the Foreman video sequence using a quantization parameter of 7 as before, and encoding a single macroblock per inter-encoded frame yields 997.5 kbits for the 150 frame video clip. The first frame which is intra-encoded has a size of 31.8 kbits leaving 965.7 kbits to inter-encode the remaining 149 frames. Therefore the average bit rate for the P-frames is 194.4 kbps ((965.7/149)*30 fps), which is used as a base case when the CBR connection capacity is set equal to the mean bit rate of the encoded video sequence.

When the Miss America video sequence is encoded using a QP of 7, and a single macroblock is intra-encoded with each frame the total number of bits is 238.9 kbits for all 150 frames. The first frame which is intra-encoded has a size of 13.6 kbits, leaving 225.3 kbits for the 149 inter-encoded frames. Therefore, the average bit rate of the inter-encoded frames is 45.4 kbps ((225.3/149)*30 fps).

7.3.2 Foreman and Miss America Results When Intra-encoded at the Macroblock Level

This section provides an in-depth discussion of the experiments run and the results obtained when the Foreman and Miss America video sequences are intra-encoded at the macroblock level. In each experiment, the first frame of the video sequence is intra-encoded, followed by 149 P-frames each having a single macroblock intra-encoded.

Tables 7.3 and 7.4 show the results obtained for the Foreman and Miss America video sequences, respectively. In each table, the first column shows the capacity of the CBR connection in bits/sec and as a multiple of the average bit rate. The connection utilization, jitter and buffer requirements are also shown in the tables. The video quality for the Foreman video sequence in these experiments is 4.356 because no frames are lost and a single macroblock is intra-encoded with each P-frame. Similarly, the video quality for the Miss America video sequence is 4.3.

The maximum possible utilization in this experiment for the Foreman video sequence is obtained if the entire 997.5 kbits pertaining to the 150 frames are transmitted in 2653 full cells, which have a payload of 47 bytes. Since the CBR connection is unused for the first

Link Cap.	Link Util.	Jitter	Buffer Reqts. (bits)	
(bits/sec)	(%)	(msec)	Max.	Mean.
194431 (1.0)	99.4	3.9	46312	29960.4
204153 (1.05)	96.8	3.1	39168	16773.9
213874 (1.1)	92.4	2.5	36088	11546.3
223596 (1.15)	88.5	1.8	34344	8236.7
233317 (1.2)	84.8	1.3	32464	6362.4
252760 (1.3)	78.4	0.5	31784	4537.3
272203 (1.4)	72.8	0.1	31784	3828.0
291646 (1.5)	68.0	0.0	31784	3355.6
388862 (2.0)	51.1	0.0	31784	2200.4

Table 7.3: Test 7 results, intra-encoding at the macroblock level using various link capacities and unlimited buffering for the Foreman video sequence.

Link Cap.	Link Util.	Jitter	Buffer Reqts. (bits)	
(bits/sec)	(%)	(msec)	Max.	Mean.
45360 (1.0)	98.4	5.5	23408	11220.1
47628 (1.05)	96.6	4.5	18896	8910.7
49896 (1.1)	94.8	3.7	14664	6731.5
52164 (1.15)	90.9	3.0	13640	4679.5
54432 (1.2)	87.2	2.4	13640	3186.4
58968 (1.3)	80.5	1.5	13640	1666.9
63504 (1.4)	74.8	0.7	13640	1362.5
68040 (1.5)	70.0	0.4	13640	1223.3
90720 (2.0)	52.6	0.0	13640	818.8

Table 7.4: Test 7 results, intra-encoding at the macroblock level using various link capacities and unlimited buffering for the Miss America video sequence.

1/30 second while the first frame is encoded, the first 17 cells are not utilized, therefore the maximum utilization possible is 2653/2670, or 99.4%. Analyzing the arrival and departure times of the frames in the buffer at the network source, all 150 frames are delayed beyond their single frame period, therefore the maximum utilization is achieved.

The maximum utilization for the Miss America video sequence is obtained if the entire 238.9 kbits pertaining to the 150 frames can be transmitted in 636 full cells. In this experiment the first 4 cells (corresponding to when the first frame is encoded) are not utilized. Therefore, the maximum utilization possible is 636/640, or 99.4%. However, because of a larger coefficient of variation among the inter-encoded frames, 0.07 compared to 0.04 (shown in Tables 7.1 and 7.2) for the Miss America and Foreman sequences respectively, the utilization of the connection is lower for the Miss America video sequence. Analysis of the simulation trace, pertaining to the Miss America clip when the CBR connection capacity is equal to the average bit rate, indicates that frame periods 44-48 and 56-60 do not fully utilize the available capacity since the frame sizes are less than the mean.

Figure 7.1 shows a plot of the jitter against the link capacity when the Foreman and Miss America video sequences are encoded using intra-encoding at the macroblock level. As shown in the figure, acceptable jitter values not exceeding 4 msec can be achieved using the mean bit rate for the Foreman video sequence while the Miss America sequence requires the CBR connection capacity to be increased by as little as 10% over the mean bit rate. Observe that the Foreman video stream has considerably lower amounts of jitter than the Miss America video stream when the available CBR connection capacity is low. Setting the CBR connection capacity equal to the mean bit rate of the inter-encoded frames causes frames 146-150 of the Foreman video sequence to be queued in the buffer after the last frame has been added. However, frame numbers 61-150 of the Miss America video sequence are delayed beyond a single frame period which results in frames 139-150 being queued after the last frame is encoded and significantly affects the resulting jitter.

Table 7.5 summarizes information about the frame sizes of the Foreman and Miss America video sequences. Columns 2-5 contain information pertaining to the P-frames which include the mean, standard deviation, coefficient of variation and peak/mean ratio. The last column shows the size of the I-frame (frame number 1) in each video sequence. As shown in the table, the coefficient of variation indicates the Miss America video frame sizes are more variable than those of the Foreman video sequence. The coefficient of variation for the Miss America P-frames is nearly 50% larger than that of the Foreman P-frames. Similarly, the peak/mean ratio, which is defined below, is larger for the Miss America inter-

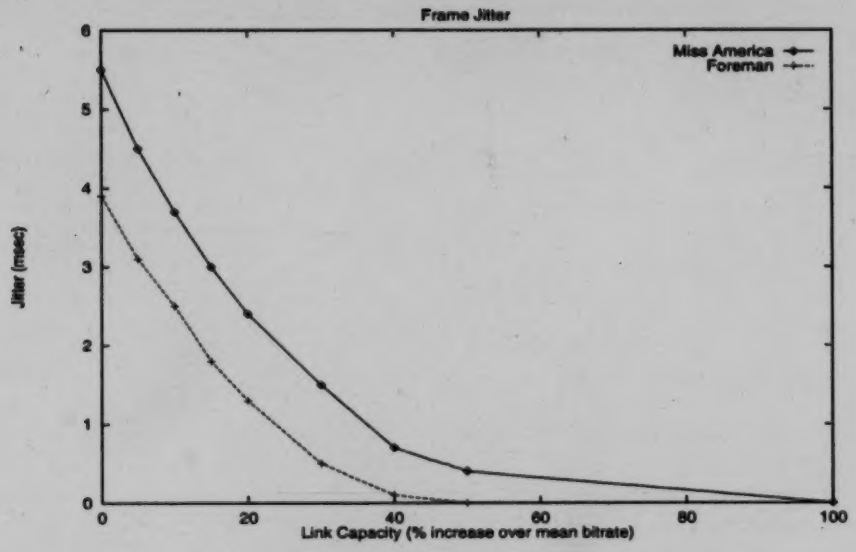

Figure 7.1: Resulting jitter values for the Miss America and Foreman video sequences using intra-encoding at the macroblock level.

encoded frames. This higher variation among the frame sizes, causes the Miss America video sequence to have a larger jitter than the Foreman video sequence.

$$\text{peak/mean ratio} = (\textit{largest P-frame})/(\textit{mean P-frame size})$$

Video Sequence	149 P-frames				I-frame
	Mean	STD	Coef. Var.	Peak/Mean	Size
Foreman	6481	1364	0.044	1.44	31784
Miss America	1512	386	0.065	1.50	13640

Table 7.5: Frame sizes in the Foreman and Miss America video sequences when a single macroblock is intra-encoded in each P-frame.

Figure 7.2 shows the maximum buffer requirements at various network link capacities for the Miss America and Foreman video sequences when a single macroblock is intra-encoded in each frame. The buffer requirements exceed the size of the large I-frame when the CBR connection capacity is low because the network is unable to transmit the first frame which is intra-encoded in a single frame period (1/30 second). Therefore, inter-encoded frames which immediately follow are also buffered.

As Figure 7.2 shows, the buffer requirements decrease to the point of being just large enough to accept the intra-encoded frame when the CBR connection capacity is increased

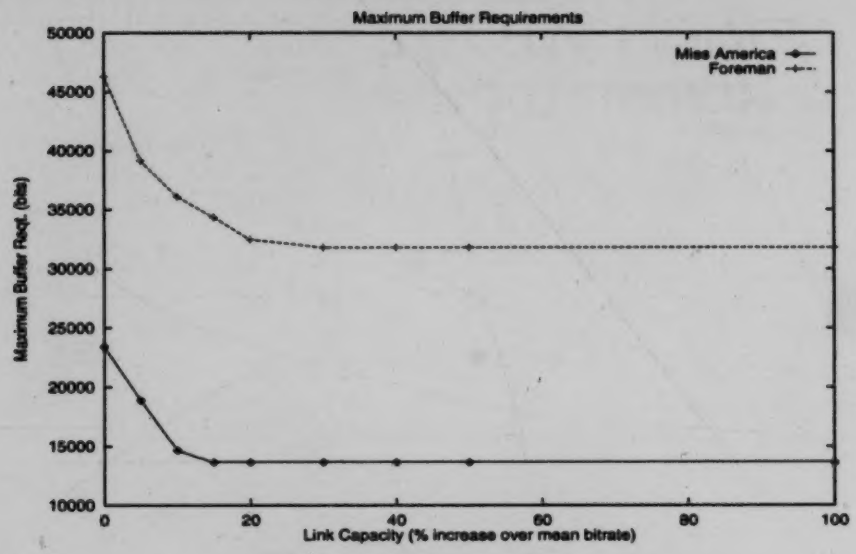

Figure 7.2: Maximum buffer requirements for the Miss America and Foreman video sequences.

by 15% over the mean bit rate for the Miss America sequence, and 30% for the Foreman video sequence. Since the frames are sent to the ATM adaptation layer as complete frames rather than a continuous bit stream, buffer requirements matching the size of the largest frame are expected. The Foreman video sequence requires a greater increase in the CBR connection capacity to reduce the buffer requirements to the size of the I-frame because the size of the first few P-frames are larger than the mean P-frame size of 6,481.0 bits. This in turn causes the buffer requirements to be larger, and a higher CBR connection capacity is required to alleviate the conditions where the buffer size required is larger than the I-frame. Since the first few frames of the Miss America video sequence are less than the mean P-frame size of 1,512 bits, increasing the CBR connection capacity by only 15% above the mean inter-encoded frame bit rate reduced the buffer requirements to the size of the I-frame.

7.4 Performance Comparison of Intra-encoding at the Frame and Macroblock Levels

This section compares the performance when intra-encoding at the frame level and macroblock level are used to encode the Foreman and Miss America video sequences. The results corresponding to intra-encoding at the frame level were obtained from Test 4 (Section

6.2.4) where a dynamic video encoder is used and the small-sized frames are discarded when buffer overflow is encountered.

Table 7.6 shows the connection capacity, utilization, and jitter for the Foreman video sequence when intra-encoding is performed at the frame level and macroblock level. The results in the table show the CBR connection capacities that are required to achieve an acceptable level of jitter; the results corresponding to higher capacity connections have been omitted. Intra-encoding the Foreman video sequence at the frame level requires the CBR connection capacity to be increased to twice the average bit rate in order to lower the jitter to an acceptable level. In addition, the corresponding utilization of the connection decreases to below 51%. However, intra-encoding the sequence at the macroblock level produces an acceptable jitter of 3.9 msec and a high connection utilization of 99.4% when the connection capacity is equal to the average bit rate. The performance of intra-encoding at the frame level is poor because every 99th frame is an I-frame, and I-frames are much larger than the average P-frame size.

Frame Level Intra-encoding			Macroblock Level Intra-encoding		
Link Cap. (bits/sec)	Link Util. (%)	Jitter (msec)	Link Cap. (bits/sec)	Link Util. (%)	Jitter (msec)
192128 (1.0)	93.8	11.6	194431 (1.0)	99.4	3.9
288192 (1.5)	67.3	6.1			
384256 (2.0)	50.5	3.8			

Table 7.6: Results obtained using the frame and macroblock level intra-encoding of the Foreman video sequence.

Table 7.7 shows the connection capacity, utilization, and jitter for the Miss America video sequence when intra-encoding is performed at the frame and macroblock levels. The results in the table indicate that in order to achieve an acceptable jitter value for the Miss America video sequence the connection capacity must be increased by four times the mean bit rate when intra-encoding is performed at the frame level, reducing the utilization to nearly 26%. However, intra-encoding at the macroblock level produces a jitter of 3.7 msec when the connection capacity is increased by only 10% over the mean bit rate and achieves a connection utilization of nearly 95%.

A comparison of the connection capacities corresponding to the mean bit rate when intra-encoding is done at the frame level and then at the macroblock level shows that the capacities are very similar. When comparing the bit rates of these two encoding schemes, it is important that a similar number of macroblocks are intra-encoded. Therefore, we compare

Frame Level Intra-encoding			Macroblock Level Intra-encoding		
Link Cap. (bits/sec)	Link Util. (%)	Jitter (msec)	Link Cap. (bits/sec)	Link Util. (%)	Jitter (msec)
44149 (1.0)	88.9	27.3	45360 (1.0)	98.4	5.5
66224 (1.5)	66.8	15.7	47628 (1.05)	96.6	4.5
88298 (2.0)	51.3	10.9	49896 (1.1)	94.8	3.7
132447 (3.0)	34.6	6.3			
176596 (4.0)	26.2	4.0			

Table 7.7: Results obtained using the frame and macroblock level intra-encoding of the Miss America video sequence.

the average bit rate corresponding to the encoding of frame numbers 2-100 inclusive. This includes 98 P-frames and a single I-frame when intra-encoding is performed at the frame level, and 99 P-frames which each have a single macroblock intra-encoded when intra-encoded at the frame level. Intra-encoding the Foreman video sequence at the frame level yields a mean bit rate of 195 kbps while intra-encoding at the frame level results in a mean bit rate of 194 kbps for 99 frames. The Miss America video sequence generates 47 kbps when each of the algorithms are used.

We have shown in this section that the results generated when the Foreman and Miss America video sequences are intra-encoded at the macroblock level; acceptable levels of jitter are achieved using lower connection capacities compared to when intra-encoding at the frame level is used. Similarly, the corresponding utilization of the CBR connection is higher when intra-encoding is performed at the macroblock level rather than the frame level. Finally, the number of bits generated by both encoding algorithms are very similar.

7.5 Results When the Carphone, Claire, and Salesman Video Sequences are Intra-encoded at the Macroblock Level

In this section, we discuss the performance of our macroblock intra-encoding algorithm using the three remaining video sequences: Carphone, Claire, and Salesman. Again, we evaluate the required CBR connection capacities to reduce the jitter to an acceptable value not exceeding 4.0 msec.

Table 7.8 shows the connection capacity, utilization, jitter, and buffer requirements for each of the video sequences used in this experiment. As shown in the table, the CBR connection capacity must be increased by 10%, 15% and 30% (over the mean bit rate of the P-frames) for the Carphone, Claire and Salesman video sequences respectively, in order to

reduce the jitter to an acceptable level. The results for the Salesman video sequence include additional entries corresponding to when the link capacities are increased to 30% since this is when the jitter is reduced to within 4.0 msec.

Video Sequence	Link Cap.	Link Util.	Jitter	Buffer Reqts. (bits)	
	(bits/sec)	(%)	(msec)	Max.	Mean.
Carphone	130410 (1.0)	99.4	5.0	70984	51142.2
"	136930 (1.05)	98.6	4.2	57472	37012.7
"	143451 (1.1)	94.0	3.5	48824	23120.3
"	149972 (1.15)	90.0	2.8	42128	15267.3
"	156492 (1.20)	86.3	2.2	35360	10656.8
Claire	42377 (1.0)	99.5	5.6	24360	17773.4
"	44496 (1.05)	99.5	5.0	21176	13413.2
"	46615 (1.1)	98.8	4.2	21176	8538.7
"	48734 (1.15)	94.1	3.7	21000	6252.6
"	50852 (1.20)	90.7	3.0	21000	5139.3
Salesman	73091 (1.0)	99.4	8.7	45336	30388.4
"	76746 (1.05)	97.7	7.7	42728	23645.1
"	80400 (1.1)	94.5	6.8	40848	19590.2
"	84055 (1.15)	91.6	5.9	38592	15808.3
"	87709 (1.20)	89.1	5.2	36712	12840.2
"	91364 (1.25)	86.1	4.5	34832	10151.9
"	95018 (1.30)	82.8	3.8	32576	8459.5

Table 7.8: A summary of the results obtained for the Carphone, Claire, and Salesman video sequences using various link capacities and unlimited buffering.

These video sequences all obtain the maximum possible utilization of the CBR connection when the capacity is equal to the mean bit rate. The buffer congestion caused by transmission of the first I-frame ensures the buffer does not empty until after the last frame has been transmitted. The maximum utilization for the Claire video sequence is 0.1% higher than the other two sequences and is achieved since the 231,472 bits are transmitted in 616 cells, an additional three cells are not utilized during the encoding of the first frame.

The Salesman video sequence requires a substantially higher increase, 30% over the mean bit rate in order to achieve a jitter not exceeding 4.0 msec, while the next contender Claire, requires only a 15% increase over the mean bit rate. When transmitting the Salesman video sequence with a CBR connection capacity equal to the mean bit rate of the P-frames, transmission of the large I-frame causes frames 2-14 to be buffered before the first frame leaves the buffer. Similarly, the buffer congestion caused by the first frame when higher CBR connection capacities are used results in higher jitter values when compared with

other video sequences.

The larger jitter values corresponding to the Salesman video sequence are explained by examining the variability of the P-frames, and the how closely sized the mean P-frame and the I-frame (the first frame encoded) are. When the ratio of the mean P-frame size (used to compute the CBR connection capacity) and the I-frame size is large, the buffer congestion and jitter caused by the first I frame is reduced. Similarly, if the variance among the P-frame sizes is low, the likelihood of overloading the CBR connection is reduced and the corresponding jitter is lower.

7.6 Summary of Results (Intra-encoding at the Macroblock Level)

This section summarizes the variances of the frames sizes pertaining to the five video sequences and how they effect the transmission performance. As shown in Table 7.9, the Salesman clip has nearly the lowest ratio when the mean P-frame and I-frame are compared, second only to the Claire sequence where the mean P-frame size is 7% of the I-frame. In addition, the Salesman has a substantially higher variability (coefficient of variation) among its P-frames, more than two times that of the next highest sequence, Claire. These two factors require the CBR connection capacity be increased by 30% over the mean bit rate in order for the Salesman video to be transmitted with a jitter not exceeding 4 msec. The Foreman video sequence which has a low variability among the inter-encoded frames, and a mean P-frame size that is a substantial 20% of the I-frame size, achieves a jitter of 3.9 msec using a CBR connection equal to the mean bit rate.

Video Sequence	149 P-frames			I-frame	Ratio
	Mean	STD	Coef. Var.	Size	(P-frame(mean)/I-frame)
Carphone	4347	1119	0.066	29384	0.15
Claire	1413	376	0.071	21000	0.07
Miss America	1512	386	0.065	13640	0.11
Foreman	6481	1364	0.044	31784	0.20
Salesman	2436	964	0.157	31944	0.08

Table 7.9: Summary of the frame sizes pertaining to the five sequences we tested when a single macroblock is intra-encoded in each P-frame.

We chose the Foreman and Miss America video sequences for testing in all of our experiments because these two sequences have the highest and lowest amount of (spatial and temporal) information present, respectively. However, as the results of this experiment il-

lustrate, the variability of the inter-encoded frames and the closeness of the mean P-frame size and the I-frame, are parameters that have a greater effect on the performance of the video transmission when utilizing a CBR connection. Video sequences having a low variance among the P-frames and a large mean P-frame size to I-frame size ratio, can be successfully transmitted using a CBR connection capacity equal to the mean bit rate of the inter-encoded frames as we have shown with the Foreman video sequence.

Chapter 8

Conclusion and Future Work

8.1 Conclusion

VBR video compression is preferred over CBR encoding because the video quality remains constant, although variable-sized compressed frames are produced. If the compressed bit stream can be transmitted over a CBR ATM connection without exceeding the specified peak rate, losses caused by congestion within the network can be avoided. In this thesis we investigate the performance when VBR video is transmitted over a CBR connection. In addition, we introduce two video encoders types, a black box and a dynamic video encoder and evaluate their performance when frames are lost due to buffer overflow. Because intra-encoding at the frame level results in poor network performance we also evaluate the performance when intra-encoding is performed at the macroblock level.

We show that the CBR option of the H.263 video encoder we are using does not provide adequate traffic smoothing of the compressed bit stream created, but actually makes it more volatile than when VBR encoding is used. We then investigate the performance of transmitting VBR encoded video over a CBR ATM connection. Traffic smoothing is achieved using a buffer implemented between the video encoder and the network source. Using the leaky bucket algorithm, data is added to the buffer at a variable rate and removed from the buffer at a constant rate determined by the CBR connection capacity.

Our experiments in Chapter 6 show that intra-encoding at the frame level results in low network utilization and high delays due to the massive size of the intra-encoded frames compared to the inter-encoded frames. Similarly, we illustrate that inter-encoding is sensitive to frame losses and should be avoided since additional errors are introduced degrading the video quality. Using the dynamic encoder that we developed, superior video quality is obtained since it encodes each frame using the frame that will be used by the decoder (assuming no losses occur in the network). Thus the video degradation introduced by the

black box encoder when frames are discarded is avoided.

Although losses of video frames should be avoided, if a VBR ATM quality of service is being used, the large inter-encoded frames and intra-encoded frames should be assigned high priority in the network. This will cause small inter-encoded frames to be discarded should congestion occur. However, this will typically result in better video quality than if the large-sized frames are discarded since the large amounts of temporal information associated with the larger inter-encoded frames is preserved.

To address the low network utilization and high delays associated with intra-encoding entire frames, in Chapter 7 we introduce intra-encoding at the macroblock level. Intra-encoding a single macroblock in each of the frames ensures every 99 frames, each of the 99 macroblocks of a video frame have been intra-encoded once, and adheres to the H.263 recommendation that macroblocks be intra-encoded at least once every 132 times. Using the mean bit rate generated by the 149 inter-encoded frames, the Foreman video sequence produces a video quality rating of 4.356 and a tolerable amount of jitter, 3.9 msec, while obtaining a remarkable 99.4% link utilization. The Miss America video sequence produces a video quality rating of 4.300, but requires the CBR connection capacity be increased by 10% above the mean bit rate in order to reduce the jitter to an acceptable amount of 3.7 msec. The link utilization is 94.8%.

For completeness, we also evaluate the performance of the other three video sequences, Carphone, Claire, and Salesman, when intra-encoding is performed at the macroblock level. The Carphone and Claire sequences require the capacity of the CBR connection be increased by 10% and 15% respectively over the mean bit rate in order to reduce the jitter to an acceptable level that does not exceed 4 msec. The Salesman video sequence which has a coefficient of variation of 0.157, substantially larger than the other sequences, requires the largest increase of the CBR connection capacity 30% over the mean bit rate of the inter-encoded frames in order to reduce the jitter to within 4.0 msec. It also has the lowest link utilization among all the video sequences, nearly 83%, once the jitter has been reduced to an acceptable level.

The number of bits produced by a video sequence when a single macroblock is intra-encoded in each frame (using QCIF) produces a similar number of bits compared to when every 99th frame is intra-encoded. However, the uniformity of the frames sizes produced substantially improves the network performance. Small increases in the CBR connection capacity over the mean bit rate are able to reduce the jitter to acceptable levels and obtain a much higher utilization of the CBR connection.

Our conclusions are largely based on the results obtained from experiments using only two video sequences, the Foreman and Miss America sequences. However, we believe that in general, intra-encoding at the macroblock level will provide acceptable levels of jitter using a lower capacity, CBR connection, compared to when intra-encoding is performed at the frame level.

8.2 Future Work

There are several issues which we intend to address as part of our future work. In our experiments we did not model the loss of partial frames, instead, the whole frame was discarded when it did not fit in the buffer space available. We plan to investigate the ability to reassemble partial frames received in order to preserve as much information as possible. In addition, our experiments use a round-robin algorithm for selecting the macroblocks to be intra-encoded rather than possibly updating more important areas such as near the center of the video frame. We also plan to investigate improving the video quality assessment algorithm used since increasing the number of macroblocks intra-encoded does not continuously increase the video quality rating. These topics are discussed in greater detail below.

Our tests do not simulate losses within the network since the capacity of the CBR connection is not exceeded. However, it would be interesting to simulate the effects when partial frames are lost within the network. A simple method for addressing the problem when part of an H.263 compressed video frame is lost (for example due to congestion within the network) is to discard the entire frame, which avoids the overhead of finding the frame boundaries and reassembling partial frames. However, in periods of high motion such as the Foreman video clip, losses of entire frames can lead to jerky motion thus degrading the video quality. Therefore, we plan to investigate the possibility of reconstructing video frames using the cells which are successfully transmitted. Depending on the cell(s) lost, this may map to a macroblock, thus the entire frame could be displayed and the missing macroblock filled using interpolation possibly from the previous frame. Although there may be a slight discontinuity within the image, this method is expected to provide better results and would achieve a higher video quality assessment rating than the method currently used. Similarly, if we are able to reconstruct partial H.263 frames, the granularity of our proposed traffic smoothing algorithm could be refined from discarding frames to possibly discarding partial frames which may correspond to the macroblock level.

Work done by Wiebe and Basu [38] indicates macroblocks corresponding to the *fovea*

(area of interest) should present more detail than the *periphery*, areas which are not part of the fovea. These ideas could also be used to determine which areas of a video frame should be intra-encoded more frequently in lossy environments in order to further improve the video quality. This is expected to provide superior results compared to when the macroblocks to be intra-encoded are chosen in a round-robin fashion.

We plan to investigate improving the video quality assessment algorithm used with respect to the amount of temporal information lost. In our experiments where more information is transmitted, such as when more macroblocks are intra-encoded, the resulting video quality does not continuously increase as shown in Tables 7.1 and 7.2. These slight decreases are due to fluctuations in the m_3 parameter which selects the frame with the largest added motion. However, this parameter is based on a ratio involving the temporal information of the original and encoded frame; this ratio does not give any emphasis on the quantity of temporal information present. Therefore, frames which contain a small amount of temporal information may be selected for the third parameter causing the overall video rating to be reduced unfairly. We hope to propose a method that will address this unfairness.

Bibliography

[1] H. Balakrishnan, V. Padmanabhan, S. Seshan, and R. Katz. A comparison of mechanisms for improving tcp performance over wireless links. *IEEE/ACM Transactions on Networking*, 5(6):756–769, December 1997.

[2] M. Baldi and Y Ofek. End-to-end delay of videoconferencing over packet switched networks. In *Infocom '98*, pages 1084–1092, San Francisco, California, March 1998.

[3] J. Beran, R. Sherman, M. Taqqu, and W. Willinger. Long-range dependence in variable-bit-rate video traffic. *IEEE Transactions on Communications*, 43(2-4):1566–1579, February 1995.

[4] Medhavi Bhatia. Qos over the internet: The RSVP protocol. *http://www.rvs.uni-hannover.de/people/neitzner/Studienarbeit/literatur/more/rsvp.html*, 1998.

[5] J. Chen and D. Lin. Optimal bit allocation for coding of video signals over ATM networks. *IEEE Journal on Selected Areas in Communications*, 15(6):1002–1015, August 1997.

[6] Dr. Leonardo Chiariglione. The mpeg home page. *http://drogo.cselt.stet.it/ufv/leonardo/mpeg/index.htm*, July 1998.

[7] Dr. Leonardo Chiariglione. Overview of the mpeg-4 standard. *http://drogo.cselt.stet.it/ufv/leonardo/mpeg/standards/mpeg-4/mpeg-4.htm*, July 1998.

[8] I. Dalgic and F. Tobagi. Performance evaluation of ATM networks carrying constant and variable bit-rate video traffic. *IEEE Journal on Selected Areas in Communications*, 15(6):1115–1131, August 1997.

[9] M.W. Garret and W. Willinger. Analysis, modeling and generation of self-similar VBR video traffic. In *SIGCOMM Symposium on Communications Architectures and Protocols*, pages 269–280, London, UK, September 1994.

[10] S. Gringeri, B. Khasnabish, A. Lewis, K. Shuaib, R. Egorov, and R. Basch. Transmission of MPEG-2 video streams over ATM. *IEEE Multimedia*, pages 58–71, January 1998.

[11] Line transmission of non-telephone signals, video coding for low bitrate communication. ITU-T, Draft H.263, May 1996.

[12] M. Ibrahim, A. Hamdi. Efficient admission control for real-time MPEG-2 VBR video sources. In *LANMAN '98*, pages 63–69, Banff, Alberta, May 1998.

[13] R. Koenen. MPEG-4 overview - (Dublin version). *http://garuda.imag.fr/MPEG4/syssite/syspub/index.html*, July 1998.

[14] F. Kossentini, Y. Lee, M. Smith, and R. Ward. Predictive rate distortion optimized motion estimation for very low bit-rate video coding. *IEEE Journal on Selected Areas in Communications*, 15(9):1752–1763, December 1997.

[15] H. Kwon, M. Venkatramam, and N. Nasrabadi. Very low bit-rate video coding using variable block-size entropy-constrained residual vector quantizers. *IEEE Journal on Selected Areas in Communications*, 15(9):1714–1725, December 1997.

[16] W. Leland, M. Taqqu, W Willinger, and D. Wilson. On the self-similar nature of Ethernet traffic (extended version). *IEEE/ACM Transactions on Networking*, 2(1):1–14, February 1994.

[17] H. Li, A. Lundmark, and R. Forchheimer. Image sequence coding at very low bitrates: A review. *IEEE Transactions on Image Processing*, 3(5):589–609, September 1994.

[18] S. Liew and C. Tse. Video aggregation: Adapting video traffic for transport over broadband networks by integrating data compression and statistical multiplexing. *IEEE Journal on Selected Areas in Communications*, 14(6):1123–1137, August 1996.

[19] S. Liew and D. Tse. A control-theoretic approach to adapting VBR compressed video for transport over a CBR communications channel. *IEEE Journal on Selected Areas in Communications*, 6(1):42–55, February 1998.

[20] M. Liou. Overview of the px64 kbit/s video coding standard. *Communications of the ACM*, 34(4):59–63, April 1991.

[21] H. Liu and M. Zarki. Performance of H.263 video transmission over wireless channels using hybrid ARQ. *IEEE Journal on Selected Areas in Communications*, 15(9):1775–1786, December 1997.

[22] W. Luo and M. Zarki. Quality control for VBR video over ATM networks. *IEEE Journal on Selected Areas in Communications*, 15(6):1029–1039, August 1997.

[23] J. McManus and K. Ross. Video-on-demand over ATM: Constant-rate transmission and transport. *IEEE Journal on Selected Areas in Communications*, 14(6):1087–1098, August 1996.

[24] S. Okubo, G. Dunstan, S. Morrison, M. Nilsson, H. Radha, D. Skran, and G. Thom. ITU-T standardization of audiovisual communication systems in ATM and LAN environments. *IEEE Journal on Selected Areas in Communications*, 15(6):965–982, August 1997.

[25] W. Olesinski and P. Gburzynski. Real-time traffic in deflection networks. In *Proceedings of Communication Networks and Distributed Systems Modeling and Simulation (CNDS'98)*, pages 23–28, San Diego, California, January 1998.

[26] G. Pacifici and G. Karlsson. Guest editorial real-time video services in multimedia networks. *IEEE Journal on Selected Areas in Communications*, 15(6):961–963, August 1997.

[27] E. Raymond. Signal-to-noise ratio. *http://pf2.phil.uni-sb.de/fun/jargon/signal-to-noise_ratio.html*, July 1996.

[28] D. Reininger, D. Raychaudhuri, and J. Hui. Bandwidth renegotiation for VBR video over ATM networks. *IEEE Journal on Selected Areas in Communications*, 14(6):1076–1086, August 1996.

[29] D. Saha, S. Mukherjee, and Tripathi S. Multirate scheduling of VBR video traffic in ATM networks. *IEEE Journal on Selected Areas in Communications*, 15(6):1132–1147, August 1997.

[30] G. Schuster and A. Katsaggelos. A theory for the optimal bit allocation between displacement vector field and displaced frame difference. *IEEE Journal on Selected Areas in Communications*, 15(9):1739–1751, December 1997.

[31] MacWEEK staff. Quicktime gets nod for mpeg-4. *http://macweek.zdnet.com/mw_1207/nw_quicktime.html*, 1998.

[32] T. Suzuki. ATM adaptation layer protocol. *IEEE Communications Magazine*, pages 80–83, 1994.

[33] A. Tanenbaum. *Computer Networks*. Prentice Hall, Upper Saddle River, New Jersey, third edition, 1996.

[34] Norway Telenor R&D. H.263 encoder source code. *ftp://bonde.nta.no/pub/tmn/software*, August 1996.

[35] T. Turletti and C. Huitema. Videoconferencing on the Internet. *IEEE/ACM Transactions on Networking*, 4(3):340–351, June 1996.

[36] D. Tzovaras, S. Vachtsevanos, and M. Strintzis. Optimization of quadtree segmentation and hybrid two-dimensional and three-dimensional motion estimation in a rate-distortion framework. *IEEE Journal on Selected Areas in Communications*, 15(9):1726–1738, December 1997.

[37] A. Webster, C. Jones, M. Pinson, S. Voran, and S. Wolf. An objective video quality assessment system based on human perception. In *Proceedings of SPIE Human Vision, Visual Processing, Digital Display TV*, pages 15–26, San Jose, California, February 1993.

[38] K.J. Wiebe and A. Basu. Improving image and video transmission quality over ATM with foveal prioritization and priority dithering. In *Proceedings of IAPR/IEEE, International Conference on Pattern Recognition*, Vienna, Austria, August 1996.

[39] M. Willebeek-LeMair and Z. Shae. Robust H.263 video coding for transmission over the Internet. In *Infocom '98*, pages 225–232, San Francisco, California, March 1998.

[40] K. Zhang, M. Bober, and J. Kittler. Image sequence coding using multiple-level segmentation and affine motion estimation. *IEEE Journal on Selected Areas in Communications*, 15(9):1704–1713, December 1997.

Appendix A

Glossary of Terms

- **Black box video encoder:** has no knowledge of the buffer occupancy and is unaware when frames are discarded due to buffer overflow. Therefore, when frames are discarded, inter-encoded frames reference different frames during encoding and decoding which introduces errors.
- **Cells:** are 53 byte ATM packets which have a header consisting of at least 5 bytes. In this thesis, cells have a 6 byte header and 47 byte payload.
- **Dynamic video encoder:** is aware when frames are discarded due to buffer overflow. Therefore, inter-encoded frames reference the same frames during encoding and decoding.
- **H.263:** is a video compression standard specified by the International Telecommunications Union - Telecommunication intended for bit rates below 64 kbps.
- **I-frames:** are intra-encoded frames and therefore exploit only the spatial redundancies present in a video frame.
- **Jitter:** is the standard deviation of the inter-frame display times. Jitter in excess of ± 4 milliseconds may lead to jerky or frozen scenes.
- **Macroblock:** is a part of the H.263 compression hierarchy and pertains to an area of 16x16 pixels.
- **P-frames:** are inter-encoded frames that exploit the spatial and temporal redundancies present in the frame. P-frames reference the previously encoded frame in order to exploit the temporal redundancies present.
- **QCIF:** Quarter Common Intermediate Format specifies the frame resolution is 144 lines by 176 pixels/line, and the frame luminance information precedes the chrominance information when stored in a file.
- **Quantization:** is an irreversible operation used to increase the compression of video frames by reducing the number of intensity levels found in a picture.
- **Self-similar traffic:** is traffic that illustrates bursty characteristics over a large range of time scales. Therefore, the burstiness of the traffic is scale invariant.
- **Traffic smoothing:** removes the burstiness found in network traffic. In this thesis, a buffer is used to smooth the VBR encoded video traffic stream which is transmitted over a CBR ATM connection.

- **Utilization:** of a connection (or buffer), is a percentage reflecting the capacity used compared to the capacity available. For example, if the average bit rate of the data transmitted is 25 kbps and the connection capacity is 50 kbps, the utilization would be 50%.
- **Video:** pertains to the displaying of frame sequences on a television set or computer screen.
- **Video compression:** exploits the spatial and temporal redundancies found in a video sequence in order to reduce the amount of data transmitted over a network or stored in a database.

Appendix B

Glossary of Acronyms

- AAL: ATM Adaptation Layer
- ATM: Asynchronous Transfer Mode
- CBR: Constant Bit Rate
- CLP: Cell Loss Priority
- CONV: Convolution Operator
- CS: Convergence Sublayer
- DCT: Discrete Cosine Transform
- FEC: Forward Error Correction
- GOB: Group Of Blocks
- GOP: Group Of Pictures
- GFC: General Flow Control
- HDTV: High Definition Television
- HEC: Header Error Check
- ITU-T: International Telecommunication Union - Telecommunication
- JPEG: Joint Photographic Experts Group
- LAN: Local Area Network
- MB: MacroBlock
- MID: Multiplexing Identification
- MPEG: Motion Picture Experts Group
- NNI: Network-Network Interface
- PMD: Physical Medium Dependent
- PSC: Picture Start Code
- PSTN: Public Switched Telephone Network

- QCIF: Quarter Common Intermediate Format
- QP: Quantization Parameter .
- RF: Radio Frequency
- RMS: Root Mean Square
- RSVP: Resource reSerVation Protocol
- SAR: Segmentation And Reassembly
- SEAL: Simple Efficient Adaptation Layer
- SI: Spatial Information
- SONET: Synchronous Optical NETworks
- SNR: Signal to Noise Ratio
- STD: STandard Deviation
- Telcos: TELephone Companies
- TC: Transmission Convergence
- TCP: Transmission Control Protocol
- TI: Temporal Information
- UDP: User Datagram Protocol
- UNI: User-Network Interface
- VBR: Variable Bit Rate
- VCI: Virtual Circuit Identifier
- VPI: Virtual Path Identifier